

**Trabajo Fin de Máster**

# **GLM Y MODELOS MIXTURADOS PARA EL ESTUDIO DE LOS EFECTOS DE LA CONSANGUINIDAD SOBRE LA FERTILIDAD EN LA DINASTÍA REAL DE LOS HABSBURGO**

**Francisco Camiña Ceballos**

**Julio 2014**

**Máster Oficial en Técnicas Estadísticas  
Universidad de Santiago de Compostela**





El presente documento recoge el Trabajo de Fin de Máster para el Máster en Técnicas Estadísticas realizado por D. Francisco Camiña Ceballos con el título “GLM y modelos mixturados para el estudio de los efectos de la consanguinidad sobre la fertilidad en la dinastía real de los Habsburgo”.

Ha sido realizado bajo la dirección de D. César Andrés Sánchez Sello que lo considera terminado y da su conformidad para la presentación y defensa del mismo.

Santiago de Compostela, a 1 de julio de 2014

Fdo. César Andrés Sánchez Sello



## Resumen

Actualmente el 10,4% de la población mundial es consanguínea. La consanguinidad, unión entre parientes, tiene efectos profundos en el genoma que se traducen en un aumento de la probabilidad de sufrir cualquier enfermedad de origen o con un componente genético. Uno de los efectos más dramáticos de las uniones entre parientes en una población es la depresión consanguínea, el descenso del valor medio de la eficacia biológica y sus dos componentes: la supervivencia pre-reproductiva y la fertilidad. Sin embargo, a diferencia de lo que sucede con los efectos de la consanguinidad sobre la supervivencia pre-reproductiva humana, que están más que probados, hay cierto desconocimiento sobre los efectos de los matrimonios entre parientes sobre la fertilidad de las parejas.

La dinastía real de la Europa de la Edad Moderna de los Habsburgo se ha probado como un magnífico laboratorio natural de consanguinidad humana. Los múltiples enlaces entre parientes, la abundante información genealógica y demográfica y la homogeneidad ambiental en las que vivieron estas dinastías a lo largo de casi 500 años de historia, hacen de esta dinastía un magnífico modelo para estudiar los efectos de la consanguinidad en el género humano.

En este trabajo de fin de máster se aplicaron modelos lineales generalizados, así como modelos mixtos para analizar el efecto de la consanguinidad de 124 parejas Habsburgo sobre su fertilidad. La fertilidad se codificó a través del número de hijos de las parejas así como del intervalo reproductivo de las mujeres. Los resultados revelan un efecto de la consanguinidad de los padres sobre la infertilidad de las parejas. Además se ha detectado un efecto de la senescencia de las mujeres sobre la depresión consanguínea.



<b>1. Introducción.</b>	
1.1. La Consaguinidad y sus implicaciones genéticas.	1
1.2. La Consanguinidad en el hombre.	3
1.3. Los efectos de la Consanguinidad.	3
1.3.1. Depresión consanguínea en supervivencia pre-reproductiva.	5
1.3.2. Depresión consanguínea en fertilidad.	5
1.4. La Consanguinidad en dinastías reales.	9
<b>2. Sujetos y Métodos.</b>	
2.1. Muestra utilizada.	13
2.2. Variables respuesta: Medidas de fertilidad.	13
2.2.1. Fertilidad Completada.	13
2.2.2. Periodo Reproductivo de las Parejas.	13
2.3. Variables explicativas: Coeficientes de consanguinidad y variables <i>nuisance</i> .	15
2.3.1. Cálculo de los coeficientes de consanguinidad.	15
2.3.2. Variables <i>nuisance</i> .	17
2.4. Diseño de los Análisis.	18
2.5. Métodos estadísticos.	18
2.5.1. Estimación de la densidad del coeficiente de consanguinidad.	20
2.5.2. Modelos Generales Linearizados (GLM).	21
2.5.2.1. Modelo de Poisson.	22
2.5.2.2. Modelos Binomial Negativo.	24
2.5.3. Modelos mixturados. Modelos de ceros inflados.	25
2.5.3.1. Modelos de Piosson.	26
2.5.3.2. Modelos Binomial Negativo.	27
<b>3. Resultados.</b>	
3.1. Estimación no paramétrica de la densidad de los coeficientes de consanguinidad.	31
3.2. Efecto de la consanguinidad sobre la fertilidad completada.	33
3.2.1. Análisis de las 124 familias en su conjunto.	33
3.2.2. Análisis de los datos correspondientes a las mujeres de 17 a 27 años de edad.	36
3.2.3. Análisis de los datos correspondientes a las mujeres de 30 a 40 años de edad.	38
3.3. Efectos de la consanguinidad sobre el periodo reproductivo corregido.	40
<b>4. Conclusiones y discusión.</b>	45
<b>5. Bibliografía.</b>	49



# ● Introducción

---



## 1.1. LA CONSANGUINIDAD Y SUS CONSECUENCIAS GENÉTICAS

La consanguinidad, del latín *consanguinitas*, es el apareamiento entre individuos emparentados. En genética de poblaciones se describe la unión entre parientes a través del coeficiente de consanguinidad ( $F$ ), que se define como la probabilidad de que un determinado individuo herede dos alelos idénticos por descendencia para un locus dado. La derivación original del coeficiente de consanguinidad fue propuesta por S. Wright (1922) a través de un análisis de correlación, mientras que la definición anterior, expresada en términos de probabilidad, es la que se usa actualmente y se debe a J. B. S. Haldane y P. Moshinsky (1939), C. W. Cotterman (1940) y G. Malécot (1948). De este modo, bajo un enfoque probabilístico se diferencian dos tipos de individuos: individuos autocigotos, aquellos que han heredado alelos idénticos por descendencia, e individuos alocigotos que portan dos alelos no idénticos; por lo tanto, según esta definición, el efecto de la consanguinidad es el incremento de la proporción de individuos que son autocigotos. Otra manera de contemplar el fenómeno de la consanguinidad es a través del coeficiente de coascendencia ( $\theta$ ). El coeficiente de coascendencia de dos individuos "X" e "Y" es la probabilidad de que cogiendo un gen al azar en el individuo X sea idéntico que un gen, del mismo locus, tomado al azar en el individuo Y (Malécot, 1966). Por lo tanto, el coeficiente de coascendencia de dos individuos X e Y es igual al coeficiente de coascendencia de su hijo Z ( $\theta_{XY} = F_Z$ ). En las poblaciones naturales, las especies vegetales son las que presentan mayor grado de consanguinidad, ya que, por ejemplo, numerosas especies de plantas son predominantemente auto-polinizadoras, es decir, se reproducen por autofecundación que es la forma más extrema de consanguinidad. En los humanos los casos más extremos de enlaces consanguíneos corresponden a las uniones incestuosas, definidas como enlaces entre parientes en primer grado: padre – hija, madre – hijo y hermano – hermana. Las descendencias de estas uniones tendrán un coeficiente de consanguinidad de 1/4 (0,25), mientras que los vástagos de un enlace entre un tío y su sobrina, entre primos hermanos y entre primos segundos tendrán coeficientes de consanguinidad de 1/8 (0,125), 1/16 (0,0625) y 1/64 (0,0156) respectivamente. La principal consecuencia de un enlace entre individuos emparentados es que su descendencia presenta un incremento de la homocigosis en relación a la descendencia de un apareamiento entre individuos no relacionados. Es decir, un gen presente en el antecesor común de los progenitores de un individuo, puede ser heredado por ambos progenitores y su descendencia puede recibirlo por partida doble siendo pues homocigota para dicho gen. Es por ello que la consecuencia más inmediata de la consanguinidad, tomada en sentido amplio en una población, es un aumento de la proporción de individuos homocigotos en comparación a una población no consanguínea (Crow y Kimura, 1970; Cavalli-Sforza y Bodmer, 1981; Falconer y Mackay, 1996).

En las poblaciones naturales la consanguinidad se puede producir debido a dos situaciones biológicas, este hecho, puede llevar a confusión en el estudio de la consanguinidad. Es más, en la naturaleza se puede dar toda una escala de grises entre estas dos situaciones puras. El primer modelo biológico de consanguinidad, característico de poblaciones con un gran número de individuos, se puede producir consanguinidad cuando los individuos se emparejan con parientes con una probabilidad mayor de la esperada por mero azar. G. Malecot denominó a este evento consanguinidad sistemática (Malecot, 1948, 1966; Crow y Kimura, 1970). En este contexto, la desviación del apareamiento aleatorio producirá que la población deje de estar en equilibrio Hardy – Weinberg, generándose un defecto de heterocigotos que puede expresarse a través del coeficiente de consanguinidad (Crow y Kimura, 1970; Falconer y Mackay, 1996; Hedrick, 2005). El segundo modelo genético de consanguinidad se refiere a los eventos o

procesos que se producen en las poblaciones finitas, de pequeño tamaño, donde se puede producir consanguinidad aunque el apareamiento sea al azar. Este segundo modelo se denomina consanguinidad panmítica o consanguinidad por deriva (Malecot, 1948; Jacquard, 1975; Templeton y Read, 1994). La deriva genética es una fuerza evolutiva que actúa de manera estocástica provocando un cambio de las frecuencias de los alelos de una generación a otra (Wright 1969; Crow y Kimura, 1970; Hedrick, 2005). En estas poblaciones se produce un aumento de la homocigosis debido a su pequeño tamaño a pesar de cumplirse la panmixia. Es más, en estos casos, lo que verdaderamente está sucediendo es que por un lado se están produciendo enlaces consanguíneos, con lo que aumenta la proporción de individuos homocigotos, y por otro lado se están perdiendo alelos por causa de la deriva, con lo que se reduce la variabilidad genética.

Otra perspectiva se abre cuando se puede disponer de las genealogías de los individuos de una población determinada, lo que en genética de poblaciones se denomina una población genealógica (Crow y Kimura 1970). En este caso particular se puede trazar el camino que tomaron los genes a lo largo de las generaciones y calcular el coeficiente de consanguinidad total de los individuos de dicha población, ya sea esta debida a una desviación de la panmixia, a tratarse de una población pequeña, o debido a una combinación de ambas posibilidades (Falconer y Mackay, 1996). En este contexto, tanto el coeficiente de consanguinidad como el de coascendencia se calculan a partir de la información genealógica disponible asumiendo neutralidad con respecto a la selección natural, de manera que las probabilidades de transmisión de los distintos genes pueden ser calculadas a partir de las proporciones mendelianas. Es más, estos dos coeficientes no son características inherentes de los individuos, sino de la información genealógica que se disponga (Jacquard, 1975; Templeton y Read, 1994; Falconer y Mackay, 1996; Hedrick, 2005). En todo caso, en la mayoría de las ocasiones es suficiente con disponer datos de las cuatro o cinco generaciones más recientes para calcular la  $F$  con cierta precisión (Ballou, 1997; Balloux et al., 2004). Esto es debido al hecho de que los eventos consanguíneos próximos son desproporcionadamente más influyentes para el coeficiente de consanguinidad de un individuo que los eventos sucedidos a más profundidad en el pedigrí. Sin embargo, en algunas genealogías complejas la consanguinidad remota o ancestral puede contribuir sustancialmente en la consanguinidad total del individuo, por lo que, en algunas ocasiones, la exploración de los pedigrís limitada a pocas generaciones puede infraestimar el grado de consanguinidad (Boyce, 1983; MacCluer et al., 1983; Alvarez et al., 2009). Por ello, en aquellos casos donde se sospeche que la consanguinidad remota pueda ser importante – debido a estar ante genealogías muy complejas – será necesario computar los coeficientes de consanguinidad a partir de pedigrís extendidos. Además, son diversas las ventajas de realizar el esfuerzo de construir pedigrís extendidos, como por ejemplo es la posibilidad de estudiar los efectos de la consanguinidad en caracteres de la eficacia biológica tales como la mortalidad pre-reproductiva en la población, o el analizar la evolución de la propia consanguinidad y sus efectos a lo largo de las generaciones (Charlesworth y Willis, 2009). Por tanto, se puede concluir que la ciencia genealógica es de vital importancia para el estudio de la consanguinidad en las poblaciones humanas (Tetushkin, 2011).

## 1.2. LA CONSANGUINIDAD EN EL HOMBRE

En la actualidad se considera que las parejas consanguíneas y sus descendientes suponen aproximadamente el 10,4% de la población mundial (Bittles y Black, 2010b; PRB., 2011). En demografía se considera un enlace consanguíneo aquel que se produce entre individuos cuyo grado de parentesco es de primos segundos o mayor. Esto es debido a que por debajo de este parentesco los efectos biológicos son indistinguibles de los observados en la población general. El enlace entre primos hermanos ( $\theta = 0,0625$ ) o entre primos segundos ( $\theta = 0,0156$ ) son los enlaces consanguíneos más comunes en el mundo (Bittles, 2012). Globalmente los matrimonios entre un tío y su sobrina ( $\theta = 0,125$ ), o primos hermanos dobles ( $\theta = 0,125$ ), son mucho menos comunes, sin embargo es posible encontrar ciertas poblaciones con alta incidencia de matrimonios entre un individuo y la hija de su hermano. En cuanto a las uniones incestuosas entre parientes de primer grado (padre – hija, madre – hijo, hermano – hermana,  $\theta = 0,25$  en los tres casos) parece existir un tabú universal contra este tipo de uniones consanguíneas. El incesto es ilegal en muchos países y está prohibido por las cinco grandes religiones, aun así, se pueden encontrar prácticas incestuosas en todas las poblaciones del globo. En todo caso, el estudio de la prevalencia del incesto en el mundo es un trabajo complicado ya que no existen registros debido a su ilegalidad y su estigma social.

## 1.3. LOS EFECTOS DE LA CONSANGUINIDAD EN EL HOMBRE

Como se ha visto los matrimonios consanguíneos aumentan, en una población, el número de personas que son portadoras de genes en homocigosis, idénticos por descendencia, en comparación a otra no consanguínea (Vogel y Motulsky, 1997; Hedrick, 2005). Las consecuencias para el organismo de un incremento de la homocigosis son claras: los caracteres genéticos recesivos se pueden presentar con mayor probabilidad en la descendencia de los enlaces endogámicos como consecuencia del incremento de homocigosis del genoma. Es decir, la consanguinidad no provoca enfermedades por sí misma, pero aumenta las probabilidades de sufrir cualquier enfermedad en la cual intervengan genes deletéreos en homocigosis. Las sociedades humanas observaron los efectos de la consanguinidad en los seres vivos al inicio de los procesos de domesticación de animales y plantas en el periodo Neolítico hace aproximadamente diez mil años. Aunque los efectos de la consanguinidad se conocen desde hace miles de años los mecanismos por los cuales actúa la consanguinidad y las causas últimas de estos efectos no se empezaron a conocer hasta el desarrollo de la genética a lo largo del siglo XX. Gracias a los conocimientos que ha proporcionado la genética los científicos han comprendido que la consanguinidad actúa aumentando el grado de homocigosis de los individuos y las poblaciones. Esto tiene un efecto profundo sobre el genoma y por lo tanto sobre un sinnúmero de enfermedades, dolencias y funciones del sistema inmunitario. La consanguinidad actúa entonces aumentando la probabilidad de sufrir enfermedades mendelianas, enfermedades complejas poligenéticas – como afecciones coronarias, cáncer, apoplejía, esquizofrenia, epilepsia y asma entre otras -, inmunodeficiencias y sepsis bacteriana (Alvarez et al., 2011).

Uno de los efectos mejor descritos y más contundentes de los apareamientos consanguíneos es el fenómeno de la depresión consanguínea. Esta se define habitualmente como la disminución de la eficacia biológica de la descendencia de matrimonios consanguíneos en una población (Charlesworth y Wills

2009). No obstante, esta definición puede ampliarse a cualquier rasgo fenotípico complejo que presente una distribución continua, y que su valor medio en la población se vea deprimido cuando se introduce consanguinidad; un ejemplo de esta definición más abierta es el reciente descubrimiento de que la altura se deprime con el aumento de la homocigosis (McQuillan et al., 2012). Que un carácter cuantitativo, como pueda ser la altura, se deprima, significa que el valor medio de dicho carácter cuantitativo en la población ha descendido de una generación a otra por causa de la consanguinidad. Debido a que la consanguinidad no cambia las frecuencias génicas sino las genotípicas, aumentando la proporción de genotipos homocigotos, los cambios en dicho valor medio deben estar conectados con la diferencia del valor genotípico entre homocigotos y heterocigotos. De teste modo, si consideramos los efectos combinados de todos los loci que intervienen en el carácter –actuando estos aditivamente- y un escenario de consanguinidad con ausencia de selección, el carácter medio de la población ( $M_F$ ) tendrá el siguiente valor:

$$M_F = \sum a(p - q) + 2 \left( \sum dpq \right) (1 - F); M_F = M_0 - 2F \sum dpq \quad (1.1)$$

Donde  $M_0$  es la media de la población panmítica ( $F=0$ ) que es:

$$M_0 = \sum a(p - q) + 2 \sum dpq \quad (1.2)$$

Donde  $+a$  es el valor genotípico del homocigoto del alelo que produce un incremento en el valor del carácter,  $-a$  es el valor genotípico del otro homocigoto,  $d$  es el valor genotípico del heterocigoto,  $F$  el coeficiente de consanguinidad,  $p$  y  $q$  las frecuencias génicas y  $M_0$  el valor medio del carácter antes de la consanguinidad (Falconer y Mackay, 1996). Según esta expresión, el cambio en el valor medio del carácter viene dado por  $-2F\sum dpq$ , que marca bajo qué circunstancias se producirá la depresión consanguínea. Uno de los requisitos más importantes que se debe dar para que se produzca la depresión de un carácter cuantitativo por causa de la consanguinidad es que el valor genotípico del heterocigoto ( $d$ ) debe ser mayor que 0. Debe haber selección direccional, esto es: que en el ejemplo de un locus con 2 alelos, el alelo que produce un incremento del valor del carácter debe dominar sobre el detrimental. Pueden darse dos situaciones posibles que cumplen este requisito; que  $d$  sea igual a  $+a$ , lo que se denomina dominancia completa, o que  $d$  sea superior a  $+a$ , lo que se denomina sobredominancia. La contribución de cada locus al valor del carácter dependerá de las frecuencias génicas ( $p$  y  $q$ ), cuando estas son intermedias ( $p = q = 0.5$ ) se producirá un mayor efecto en el cambio de la media del carácter. Otra conclusión que se puede obtener de la expresión anterior, es que cuando el efecto de los loci en el carácter es aditivo, el cambio de la media del carácter en la población por causa de la consanguinidad debe ser proporcional al grado de esta. En otras palabras, el cambio en la media del carácter debe seguir una relación lineal con el coeficiente de consanguinidad. La falta de dicha linealidad puede ser indicativo de procesos subyacentes como eventos epistáticos, o incluso, si se analiza una población a lo largo del tiempo, de la acción de la selección natural. En cuanto a las dos hipótesis que explican cómo se produce la depresión consanguínea (dominancia y sobredominancia) estudios empíricos realizados en *Drosophila melanogaster* sugieren que la depresión consanguínea está causada predominantemente por alelos deletéreos originados por la mutación y mantenidos a frecuencia muy baja en la población por la selección natural. Aun así, no se puede descartar la intervención de alelos mantenidos en la población por alguna forma de selección equilibradora como ventaja del heterocigoto o selección dependiente de las frecuencias. (Charlesworth y Charlesworth, 1999; Charlesworth y Wills, 2009).

### 1.3.1. Depresión consanguínea en supervivencia pre-reproductiva

La relación entre la consanguinidad y la reducción de la eficacia biológica, como puede ser mortalidad pre-reproductiva, en el hombre se comenzó a estudiar en plena era Victoriana, a mediados del siglo XIX. Sin embargo, es con el nuevo siglo es cuando popes como Fisher, Wright, Haldane, Cotterman o Malecot ponen las bases teóricas del estudio de la consanguinidad. Gracias a estos gigantes en la segunda mitad del siglo los estudios evolucionan (Sutter y Tabah, 1952) y se desarrollan métodos específicos para estudiar la depresión consanguínea. De esta forma en 1956 Morton Crow Y Muller proponen un modelo genético que permite estudiar y medir el efecto de la consanguinidad en la supervivencia pre-reproductiva a través del coste o lastre de consanguinidad (inbreeding load), definido como la pendiente de regresión de la supervivencia, en escala logarítmica, frente a el coeficiente de consanguinidad ( $F$ ) (Morton et al., 1956). Además este modelo sirve para estimar otros parámetros, ya que en un principio fue ideado para resolver la antigua controversia de que si la variabilidad genética era mantenida por genes deletéreos en equilibrio mutación – selección o por la selección equilibradora y alelos sobredominantes. Actualmente existen numerosas evidencias de los efectos perniciosos de la consanguinidad en la supervivencia antes de llegar a la edad adulta (25 – 60 años). La mayor parte de estas evidencias empíricas provienen de estudios con datos de mortalidad de descendientes de primos hermanos, puesto que este es el tipo de enlace más común en las poblaciones humanas. En una revisión realizada agrupando datos provenientes de 38 poblaciones localizadas en el este y sur de Asia, Oriente Próximo, África, Europa y Sudamérica, la descendencia de primos hermanos presentó un incremento medio absoluto de la mortalidad desde el nacimiento hasta los 10 años de un  $4.4\% \pm 4.6$ , este incremento absoluto varió desde prácticamente el 0% hasta el 19% (Bittles y Neel, 1994). Este dato medio fue revisado a la baja analizando 69 poblaciones residentes en 15 países de cuatro continentes, en este nuevo análisis, se determino que el incremento absoluto de la mortalidad hasta los 10 años era del 3.5% para la descendencia de primos hermanos (Bittles y Black 2010a). En otro estudio analizando 48 poblaciones, los mismos autores, determinaron que el incremento absoluto de la mortalidad infantil (0 a 1 año de vida) entre la progenie con un coeficiente de consanguinidad ( $F$ ) de 0.0625 (primos hermanos) se encuentra en torno el 1.1 al 1.3% (Bittles y Black 2010b; Bittles 2012). Sin embargo la variabilidad en estas poblaciones es muy elevada, ya que, por ejemplo, en el incremento de la mortalidad hasta los 10 años el rango de variación va desde el 0% hasta el 61.7%, como en la población urbana de Lahore (Pakistán) (Bittles, 2012). Lo mismo sucede con el incremento de la mortalidad infantil donde hay poblaciones con un incremento del 61.7% como en Yoruba (Nigeria) (Bittles, 2012). A diferencia de lo que sucede para consanguinidades moderadas como la de primos hermanos ( $F = 0.0625$ ), prácticamente se desconoce el comportamiento de la depresión consanguínea con niveles altos de la consanguinidad. El análisis de los efectos de la consanguinidad en supervivencia pre-reproductiva en las dinastías reales de la Europa de la Edad Moderna ha permitido el estudio de estos efectos en un contexto de una población con alta consanguinidad a lo largo de un periodo muy largo (Ceballos y Álvarez, 2013).

### 1.3.2. Depresión consanguínea en fertilidad

Los efectos de la consanguinidad sobre la fertilidad – el otro componente de la eficacia biológica – a diferencia de lo que sucede con la supervivencia pre-reproductiva, no son muy conocidos. Es más, los estudios que tratan este tema son pocos y sus conclusiones son muy dispares, con lo que existe controversia científica para con este tema (Robert et al., 2009). Una posible explicación de este hecho es que los estudios que tratan este asunto tienen que lidiar con ciertas limitaciones asociadas a la falta de

controles de importantes variables nuisance como la edad del matrimonio, desarrollo cultural, uso de anticonceptivos, duración del matrimonio y otros datos sociodemográficos (Rodgers et al., 2001). Por otra parte, la fertilidad, como parte de la eficacia biológica, es una variable con una arquitectura genética mucho más compleja que la supervivencia reproductiva. Incluso, Sir Ronald Fisher, en el siglo pasado, a través de su teorema fundamental de la selección natural llegó a formular que, con el tiempo, el acervo genético de los individuos con alta fertilidad dominaría sobre el acervo genético de la descendencia de los individuos con baja fertilidad (Fisher, 1930). De cumplirse este proceso, en el límite, el resultado sería una ausencia casi completa de variación genética en aquellos rasgos relacionados con la fertilidad. A raíz de los trabajos de Fisher, durante casi 50 años, se consideró que la fertilidad no tendría varianza genética y por lo tanto que este carácter tendría una baja heredabilidad. Así mismo, en ausencia de varianza genética dominante no se podría producir el fenómeno de depresión consanguínea. Actualmente, a través de estudios empíricos, se han observado variaciones de la fertilidad entre individuos y poblaciones que no sólo se deben a causas ambientales, sino que también se puede atribuir a influencias genéticas. En realidad Fisher no estaba confundido, no obstante, su teorema tiene validez en un contexto específico, teniendo en cuenta ciertas asunciones. El teorema se aplica sólo a la eficacia biológica total y no necesariamente a sus componentes individuales la supervivencia y la fertilidad (Edwards, 1994). Por otra parte, Fisher, no tuvo en cuenta “fuerzas perturbadoras” (Houle, 1992) como pueden ser la mutación o cambios en el medio natural, ya sean sociales o de otro tipo. En la especie humana en particular, a través del estudio de la correlación inter-generacional del tamaño de familia – es decir, la heredabilidad del número de hijos – en muestras y poblaciones homogéneas, se han encontrado evidencias que apuntan a que las diferencias en la fertilidad no sólo se deben a factores sociológicos, sino también a factores genéticos (Rodgers et al., 2001; Pluzhnikov et al., 2007). A pesar de este descubrimiento, se desconoce en gran medida como es la arquitectura genética de la fertilidad en el humano. De esta manera, se desconoce qué proporción de varianza genética es dominante o aditiva y además el peso que tiene en este rasgo de la pareja la mujer, el hombre o incluso la propia descendencia. Es por ello que el estudio de los efectos de la consanguinidad sobre la fertilidad de las parejas puede ser sumamente interesante, no sólo para conocer mejor los efectos de las relaciones entre parientes, también para descubrir cómo es la arquitectura genética de este parámetro de la eficacia biológica.

Los estudios publicados acerca de los efectos de la consanguinidad sobre la fertilidad han seguido distintas estrategias y sus resultados son contradictorios. Una aproximación al problema es investigar el efecto de la coascendencia de las parejas – igual al coeficiente de consanguinidad de la descendencia – sobre su propia fertilidad. De esta manera, se pretende analizar el efecto de la coascendencia de las parejas comparando la fertilidad de padres no emparentados y uniones consanguíneas. Las conclusiones que arrojan estos estudios son dispares; mientras que en algunos casos se encontró una relación positiva entre la coascendencia y la fertilidad (Bittles, et al., 2002; Helgason, et al., 2008), en otros no se encontró relación estadística alguna (Bittles et al., 2002). Bittles et al., en 2002, utilizaron datos de 30 poblaciones de países como India, Pakistán, Japón, Kuwait y Turquía para comparar el número de nacimientos producidos por parejas no consanguíneas y por parejas emparentadas agrupadas en cuatro categorías: tío – sobrina ( $\theta=0.125$ ), primos hermanos ( $\theta=0.0625$ ), primos segundos dobles ( $\theta=0.0313$ ) y primos segundos ( $\theta=0.0156$ ). Los investigadores, utilizando modelos de regresión lineal, encontraron una asociación positiva entre los cuatro niveles de coascendencia y el número de hijos por familia. Además, utilizando un test del signo, encontraron que esta asociación positiva fue estadísticamente significativa en el caso de los enlaces entre primos hermanos que tuvieron, de media, 0.26 hijos más. Uno de los

problemas de este meta-análisis es la gran heterogeneidad ambiental que presentan los datos, que podría estar produciendo artefactos estadísticos. Para evitar este problema, Bittles et al. (2002) compararon familias de primos hermanos con familias no relacionadas utilizando datos de la Encuesta Nacional de Salud y Familia de la India, llevada a cabo entre 1992 y 1993. Este archivo, además de ser una fuente de información de variables biológicas, aporta información sobre un gran número de variables sociodemográficas de gran interés. Utilizando el análisis de clasificación múltiple (MCA en sus siglas en inglés) los investigadores encontraron que la fertilidad está fuertemente influenciada por un gran número de factores como el analfabetismo, edad temprana de matrimonio, duración del matrimonio, compensación reproductiva y uso de anticonceptivos. Además, se encontró que estas variables estaban también asociadas positivamente a la consanguinidad. De este modo, cuando se descontaron los efectos de estas variables sociodemográficas en el análisis multivariante no se detectaron diferencias entre el número de hijos de parejas no emparentadas y parejas de primos hermanos. En contraste con estos estudios basados en extensas bases de datos y con resultados poco clarificadores, otros estudios aportan evidencias de un efecto positivo de la coascendencia de las parejas y su propia fertilidad en ciertas poblaciones aisladas. De esta forma, se detectó una asociación positiva entre la coascendencia y la fertilidad en las parejas de la población islandesa de individuos nacidos entre 1800 y 1965 (Helgenson et al., 2008). Islandia es una de las sociedades más homogéneas socioculturalmente y además se caracteriza por tener niveles de consanguinidad relativamente bajos. Los coeficientes de coascendencia se calcularon a partir de genealogías con 10 generaciones de profundidad, por lo que se obtuvieron datos continuos de coascendencias. Las conclusiones obtenidas por Helgason et al. 2008 fueron matizadas por Labouriau y Amorin (2008) utilizando datos del número de hijos de mujeres con edades comprendidas entre los 15 y 45 años, nacidas en Dinamarca en 1945. El cálculo de la coascendencia de las parejas danesas se realizó a partir del radio "marital" – la distancia entre el lugar de nacimiento de los conyugues -. Según la teoría clásica de la estructura genética poblacional espacial de Malécot (1959) (Kimura y Weiss, 1964; Morton, 1977) la coascendencia presenta una relación lineal y decreciente en función del radio marital. A diferencia de lo que sucedía en la muestra islandesa, en las parejas danesas se encontró una asociación negativa estadísticamente significativa entre la coascendencia y la fertilidad (Labouriau y Amorin, 2008). Aunque estos dos estudios parecen contradictorios, un análisis más fino de los dos grupos de datos muestra un comportamiento complejo de las dos variables, fertilidad y consanguinidad. Los dos estudios sugieren que la fertilidad humana, a menudo, crece con la coascendencia hasta un máximo y posteriormente decrece si la consanguinidad aumenta. Esta mezcla entre depresión híbrida (outbreeding depression) y depresión consanguínea (inbreeding depression) presenta muchas similitudes con simulaciones realizadas bajo escenarios genéticos generales donde estas dos depresiones fenotípicas de origen genético actúan simultáneamente (Schierup y Christiansen, 1996). Estos resultados muestran lo complicada que es la estructura genética de la fertilidad.

Otra aproximación que se ha seguido para estudiar la depresión consanguínea en fertilidad ha sido analizar los efectos de los coeficientes de consanguinidad de los progenitores sobre la fertilidad de la pareja. Esta aproximación fue utilizada con éxito en modelos animales así, por ejemplo, estudios realizados en roedores documentaron un efecto de la consanguinidad de las hembras en el número de crías que producían con machos no emparentados (Margulis, 1996). De esta manera, el descenso de la fertilidad se atribuyó a los efectos de los genes deletéreos de la madre consanguínea per se, y no de la descendencia. En otras especies como el guepardo, la gacela de Cuvier o el lobo gris mejicano se encontraron efectos adversos de la consanguinidad de los machos en la calidad de su propio semen

(Roldan et al., 1998; Asa et al., 2007). En los humanos, Ober et al. (1999) encontraron un efecto significativo del coeficiente de consanguinidad de las madres sobre su propia fecundidad. El estudio se llevo a cabo a lo largo de 15 años y se basó en datos de colonias Huteritas del estado americano de Dakota del Sur. Las distintas colonias de Huteritas presentan condiciones socio-económicas bastante uniformes y además comparten un mismo modo de vida, por lo que se posibilita el estudio de los efectos de la consanguinidad sin tener que lidiar con la heterogeneidad de variables socio-culturales. Los investigadores analizaron el efecto del coeficiente de consanguinidad de los padres y las madres en el intervalo entre partos que fue considerado como un tiempo de “supervivencia” ya que esta variable indirecta de la fecundidad y fertilidad puede considerarse un dato censurado de tiempo hasta el evento nacimiento. El modelo de riesgo proporcional de Cox mostró que las mujeres con un coeficiente de consanguinidad superior a 0.04 ( $F \geq 0.04$ ) tuvieron una menor fecundidad evidenciada por intervalos entre partos más largos. No se encontraron efectos significativos del coeficiente de consanguinidad de los padres. En contraste, no se encontró un efecto de los coeficientes de consanguinidad en el tamaño total de las familias para aquellas mujeres nacidas después de 1920 a pesar de que los efectos de la consanguinidad en la fecundidad de ese grupo de mujeres eran evidentes. Los resultados sugieren una compensación reproductora en aquellas mujeres más consanguíneas y menos fecundas, probablemente para poder alcanzar el tamaño de familia socialmente aceptado. Este efecto adverso del coeficiente de consanguinidad de las madres sobre la fecundidad se encontró también en una población aislada de los Alpes suizos donde los factores socio-económicos y culturales también son bastante homogéneos (Postma et al., 2010). Esta vez se analizó el efecto de los coeficientes de consanguinidad de los progenitores y la descendencia sobre el tamaño de 465 familias mediante un análisis de regresión lineal múltiple. En este análisis no se incluyeron aquellas parejas que no tuvieron hijos. Los resultados mostraron un efecto significativo del coeficiente de consanguinidad de las madres sobre el número de hijos que dieron a luz, es decir, las madres más consanguíneas tuvieron menos hijos. No se detecto efecto significativo de los coeficientes de consanguinidad de los padres o la descendencia.

El estudio de los efectos de la consanguinidad en la fertilidad o fecundidad puede complicarse añadiendo un elemento más, un grado de libertad más, la senescencia. La senescencia, en relación a los sistemas naturales, se refiere a los cambios relacionales entre los elementos del sistema por el paso del tiempo. La fertilidad es un parámetro de la eficacia biológica muy dependiente del tiempo y que muestra una disminución muy acusada con el paso de los años en muchas especies de animales y en el humano (Ricklefs et al., 2003; Broekmans et al., 2004; Nichols et al., 2005). Es más, la sensibilidad de la fertilidad a la consanguinidad puede que dependa también de la edad de los progenitores. Varios estudios empíricos han mostrado que las mutaciones pueden tener un efecto en los componentes de la eficacia biológica a ciertas edades en particular (Houle et al., 1994; Hughes et al., 2004); estos estudios están en comunión con las teorías genético evolutivas de la senescencia (Charlesworth, 1994). Las bases evolutivas de la senescencia se pueden explicar por dos teorías ampliamente aceptadas y no excluyentes. La teoría de la pleiotropía antagonista atribuye la senescencia a la fijación de alelos con efectos pleitrópicos que tienen efectos favorables en las primeras etapas de la vida, pero negativos en las últimas etapas (Williams, 1957). En contraste, la hipótesis de la acumulación de mutaciones atribuye la senescencia a la acumulación de mutaciones deletéreas con efectos tardíos en la eficacia biológica. Bajo esta hipótesis,, una nueva mutación que reduzca la fertilidad o la supervivencia pero que se exprese – y por lo tanto tenga efecto – sólo en una etapa tardía experimentará poca presión selectiva, por lo que su frecuencia en la población podría incrementarse por deriva genética (Promislow y Tatar, 1998). La misma mutación, si se

expresase en una etapa más temprana podría ser eliminada por la selección natural. Como consecuencia, es más probable que se acumulen mutaciones con expresión tardía en vez de temprana. El modelo de acumulación de mutaciones condujo a la predicción de que la depresión consanguínea debería incrementarse con la edad. Esta predicción, al menos en la *Drosophila melanogaster*, se ha cumplido empíricamente (Snook y Promislow, 2003). En el género humano, pocos son los estudios publicados que exploran la relación entre la senescencia y la depresión consanguínea en fertilidad. Robert et al. (2009) analizaron la relación entre los coeficientes de consanguinidad de mujeres, hombres, así como de la descendencia, y distintas variables de fertilidad y fecundidad en una población del noreste de Quebec mediante modelos lineales generalizados (GLM en sus siglas en inglés). Sólo se incluyeron familias con más de 3 hijos. Los resultados mostraron que el coeficiente de consanguinidad de los padres afectó significativamente a la tasa reproductiva a lo largo del periodo fértil. Los padres (varones) más consanguíneos mostraron una fuerte asimetría en el número de hijos producidos en la primera mitad del periodo reproductivo en comparación a la segunda mitad. Estos resultados sugieren que en los humanos es importante tener en cuenta aspectos temporales en el estudio de la depresión consanguínea en fertilidad.

#### 1.4. LA CONSANGUINIDAD EN DINASTIAS REALES

En los humanos, los casos más extremos de consanguinidad se encuentran en las dinastías reales y en los linajes nobles. Esto es debido principalmente a dos causas. Por un lado, la práctica común entre los miembros de las distintas casas reales de desposarse entre ellos como estrategia diplomática por razones de estado. Esta práctica convierte a estas dinastías en una población, de pequeño tamaño, más o menos cerrada, donde no hubo muchas oportunidades de introducir sangre nueva. Por otro lado, en muchas dinastías reales los matrimonios consanguíneos, como el incestuoso, fueron muy comunes e incluso favorecidos, ya que permitían minimizar la influencia de otros linajes en los asuntos reales. Era una manera de mantener pura la dinastía, que en muchas ocasiones era considerada de origen divino. Desafortunadamente intentar estudiar los posibles efectos que puede causar la alta consanguinidad en las poblaciones humanas a través de las dinastías reales de la antigüedad presenta, principalmente, dos importantes limitaciones. La primera limitación es la ausencia de información genealógica fiable con la que se pueda construir pedigrís para poder calcular los coeficientes de consanguinidad de los emperadores y reyes. La segunda limitación que presentan estas dinastías es que no se puede disponer de registros adecuados que contengan datos demográficos sobre los que analizar el efecto de la alta consanguinidad. Una manera de sortear estas limitaciones, ha sido posible gracias al avance de las técnicas en genética molecular. En un reciente estudio se consiguió inferir cinco generaciones de la 18ª dinastía egipcia, incluyendo Akhenaton y Tutankhamon, a través del análisis genético por un número de loci microsatélites de material genético de once momias reales (Hawass et al., 2010). De todas formas, hacen falta más estudios genéticos moleculares para poder obtener conclusiones de los efectos de la consanguinidad en estas dinastías.

A diferencia de las dinastías reales de la antigüedad, las dinastías de la Europa de la Edad Moderna son perfectos laboratorios para el estudio de la consanguinidad en el humano (Ceballos y Álvarez, 2013). Varias son las razones que soportan esta afirmación. En primer lugar, se puede disponer de abundante información, tanto de origen genealógico como biológico, de estas dinastías, lo que permite calcular con mucha exactitud los distintos coeficientes de consanguinidad. Así mismo, se pueden obtener distintas

medidas de la eficacia biológica, ya sea en su componente de supervivencia pre-reproductiva como de fertilidad. En segundo lugar, las dinastías reales de la Europa de la Edad Moderna son altamente consanguíneas. En estas dinastías los enlaces entre parientes cercanos fueron comunes en todas las generaciones, con lo que los niveles de consanguinidad llegaron a ser excepcionalmente altos, incluso mayores que un enlace incestuoso entre hermanos o entre padres e hijos. En tercer lugar, el análisis de los efectos de la consanguinidad en las dinastías reales de la Europa de la Edad Moderna permite realizar estudios con una profundidad de casi cinco siglos. Evidencias empíricas y bibliográficas han demostrado que esto es posible ya que a lo largo de todo este periodo no se han encontrado heterogeneidad ambiental en ninguno de los parámetros analizados en las familias reales (Peller, 1965; Kümmel, 1990; Ceballos y Álvarez, 2013).

En el Trabajo de Fin de Máster que se presenta a continuación (TFM) se va a utilizar la dinastía de los Habsburgo como laboratorio natural para analizar los efectos de la consanguinidad en su fertilidad. De esta forma, a través de este laboratorio natural de consanguinidad humana se pretende aumentar el conocimiento del efecto de los matrimonios entre parientes en la fertilidad humana.

## ● **Sujetos y Métodos**



## 2. SUJETOS Y MÉTODOS

### 2.1 Muestra utilizada.

El objetivo principal de este estudio es evaluar el efecto de la alta consanguinidad de los Habsburgo en su propia fertilidad. Para ello se han tomado datos de 124 familias desde la segunda mitad del siglo XV hasta el siglo XVIII. Para evaluar el efecto de los enlaces entre parientes en la dinastía Habsburgo se han utilizado dos medidas distintas de fertilidad: el número de hijos por pareja (fertilidad completada) y el periodo reproductivo de la pareja. Sobre estas dos variables respuesta se va a analizar el efecto de ciertas variables explicativas – coeficientes de consanguinidad de los progenitores y la descendencia – teniendo en cuenta también diversas variables nuisance.

### 2.2 Variables respuesta: Medidas de fertilidad

A partir de distintas fuentes históricas (Junceda, 1991; Gargantilla, 2005; Encyclopaedia Britannica, 2007; Geneall Project, 2008) se obtuvieron los datos necesarios para analizar el efecto de la consanguinidad. La fertilidad de los Habsburgo se codificó a través del número de hijos por pareja “fertilidad completada” así como en el periodo reproductivo de las parejas.

#### 2.2.1 Fertilidad Completada

La fertilidad completada se refiere al número de hijos por pareja. Para toda la dinastía a lo largo de 400 años de historia, en total, se produjeron 588 partos en 124 familias, con una media de 4.7 hijos por familia y una varianza de 22. Esta varianza tan alta, 4.7 veces mayor que la media, indica una sobre-dispersión en el número de hijos por pareja en la dinastía de los Habsburgo respecto de un modelo de Poisson. El número de hijos de un apareja en el género humano, debido a que es una clásica variable de recuento se distribuirá bajo una Poisson, sin embargo, la sobre-dispersión que muestran los datos sugieren que en este caso la distribución subyacente a los datos podría ser una binomial negativa. Estas posibilidades en cuanto a la distribución que subyace al número de hijos de las parejas humanas han sido tratadas por distintos autores (Brass, 1958; Imaizumi y Nei, 1970; Wood, 1994). En la figura 2.1 se presenta la distribución del número de hijos de las 124 parejas de la dinastía bajo estudio junto con la distribución esperada si esta variable siguiese una distribución de poisson o binomial negativa. En la figura 2.1 destaca la gran cantidad de ceros, por lo que es interesante resaltar que el 28.6% de las familias Habsburgo no tuvieron descendencia.

#### 2.2.2 Periodo reproductivo de las parejas

El periodo reproductivo de las parejas se ha definido como el tiempo, en años, transcurrido entre el nacimiento del primer y último descendiente de cada pareja. Para poder obtener una medida de la variable en todas las familias se definió el periodo reproductivo corregido (PRC) como el periodo reproductivo + 1. De esta manera la pareja con sólo un descendiente tendría un periodo reproductivo corregido de 1, y las parejas sin descendencia tendrán un periodo reproductivo corregido de 0. En la figura 2.2 se presenta la distribución del periodo reproductivo corregido (PRC) de las 124 familias estudiadas junto con las distribuciones esperadas para variables discretas contables: poisson y binomial negativa. Para esta variable de fertilidad sólo se consideró el análisis agrupando todas las familias. La

media del PRC en la dinastía Habsburgo fue de 6.29 años con una varianza de 47.64; 7.5 veces más grande que la media, con lo que en esta variable respuesta de fertilidad también se encuentra sobre-dispersión. Al igual que en la fertilidad completada, el 28.6% de las familias tuvieron un periodo reproductivo corregido de 0 años.

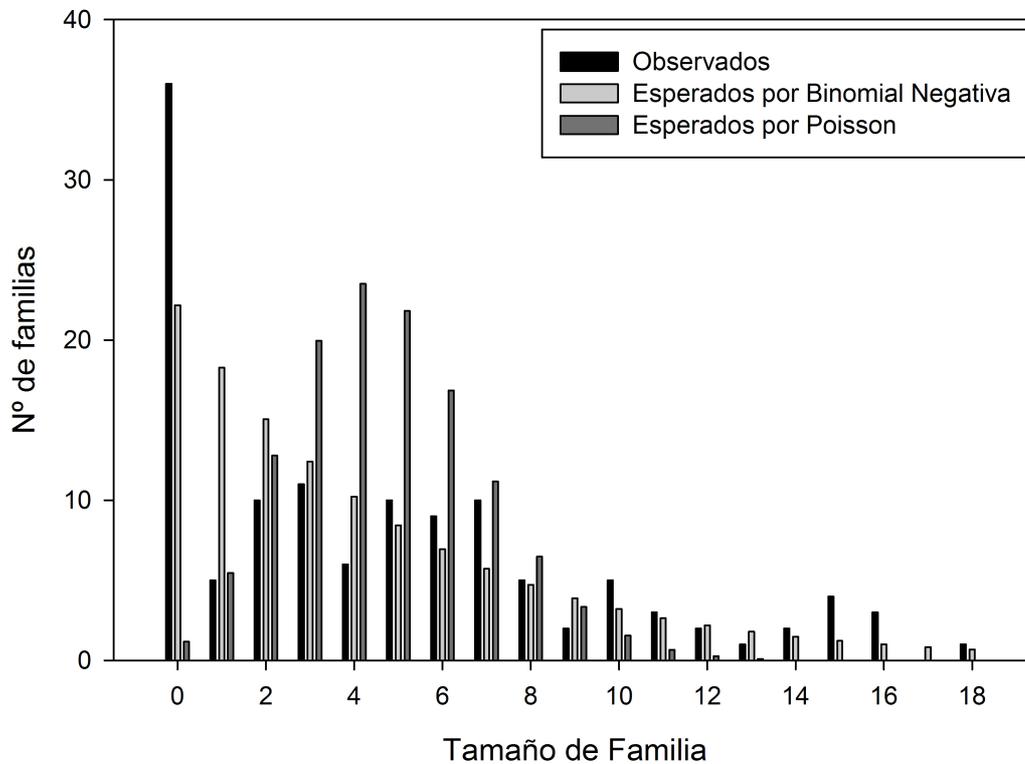


Figura 2.1. Distribución del número de hijos de las 124 parejas de los Habsburgo junto con la distribución esperada si esta variable de fertilidad siguiese una distribución de Poisson o binomial negativa.

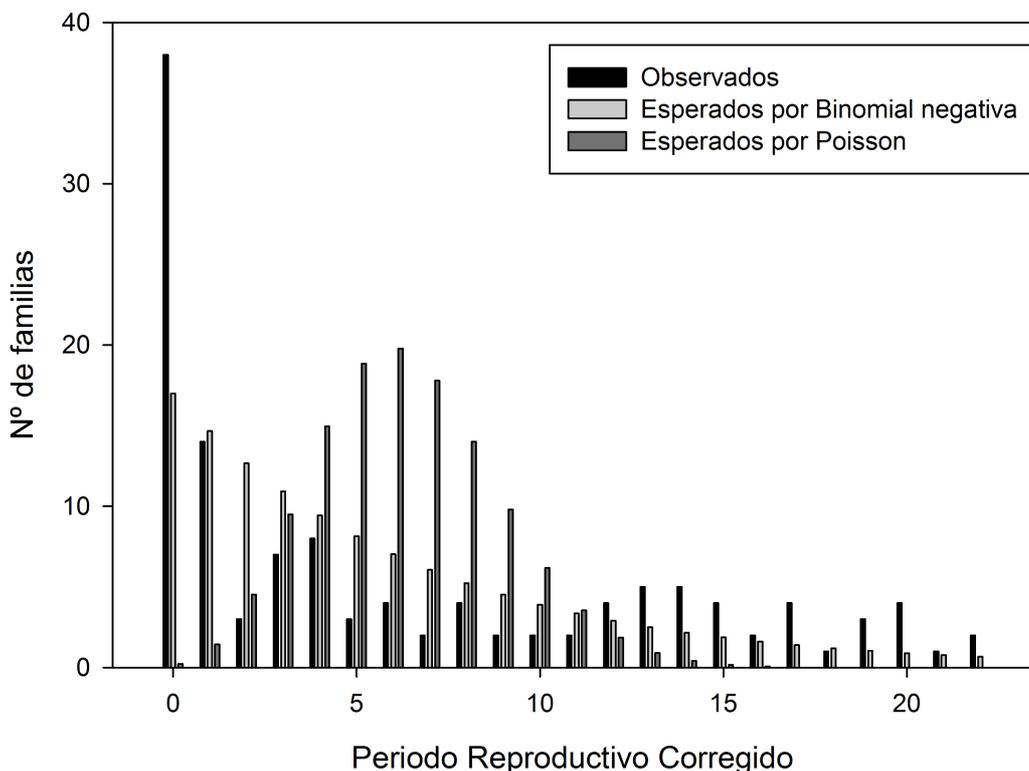


Figura 2.2. Distribución del periodo reproductivo corregido de las 124 parejas de los Habsburgo junto con la distribución esperada si esta variable de fertilidad siguiese una distribución de poisson o binomial negativa.

### 2.3 Variables Explicativas: Coeficientes de consanguinidad y variables nuisance.

Para poder analizar si la fertilidad de los Habsburgo, medida a través de la fertilidad completada y del PRC, se vio afectada por la gran cantidad de enlaces entre parientes, que tuvieron lugar en la dinastía a lo largo de 400 años, es necesario obtener los coeficientes de consanguinidad de los progenitores y descendencias. También es necesario añadir al modelo una serie de variables nuisance que tienen gran importancia en la fertilidad de las parejas Habsburgo. En la tabla 2.1 se presenta un resumen de todas las variables.

#### 2.3.1 Cálculo de los coeficientes de consanguinidad

El coeficiente de consanguinidad de un individuo es la probabilidad de heredar dos alelos idénticos por descendencia y la expresión matemática para el cálculo del coeficiente de un individuo X,  $F_X$ , a partir de su genealogía (Wright, 1922) es:

$$F_X = \sum_{i=1}^m \left(\frac{1}{2}\right)^{N_i} (1 + F_{CA(i)}) \quad (2.1)$$

Donde  $m$  es el número total de ancestros comunes a los dos progenitores de X;  $i$  es el ancestro común considerado en cada sumatorio;  $N_i$  es el número de eslabones de la cadena hasta el antepasado  $i$ , y  $F_{CA(i)}$  es el coeficiente de consanguinidad del ancestro común  $i$ . La cadena para un ancestro común  $i$  comienza en un progenitor del individuo consanguíneo, asciende hasta el antepasado común y desciende hasta el otro progenitor del individuo consanguíneo. Sin embargo, en una base de datos con más de 7000 la utilización de la fórmula derivada por Wright no es útil. Es por ello que el cálculo del coeficiente de consanguinidad se llevó a cabo mediante el programa informático comercial FSpeed (FSpeed Pro, 2005). Además del coeficiente de consanguinidad de un individuo, también es de importancia el coeficiente de coascendencia de dos individuos X e Y ( $\theta_{XY}$ ) que se define como la probabilidad de que, cogiendo un alelo al azar en el individuo X, éste sea idéntico que un alelo, del mismo locus, tomado al azar en el individuo Y (Malécot, 1966). Por lo tanto el coeficiente de coascendencia de dos individuos X e Y es igual al coeficiente de consanguinidad de su hijo Z ( $\theta_{XY} = F_Z$ ).

El cálculo de los coeficientes de consanguinidad ( $F$ ) y coascendencia ( $\theta$ ) de los individuos de la dinastía se llevó a cabo a partir de la información genealógica de los reyes y emperadores de Austrias, así como de sus hijos y nietos, se obtuvieron de distintas fuentes (Encyclopaedia Britannica, 2007; Lynch, 1991, 1992; Álvarez, 2002; Geneall Project, 2008). Esta información se incluyó en una base de datos que actualmente consta de más de 7000 individuos a lo largo de 20 generaciones partiendo desde el año 1800 y profundizando hasta el año 1100. La información genealógica incluye diferentes casas reales europeas como los Habsburgo, los Avis portugueses, los Valois y Borbones de Francia y las principales familias de la alta nobleza como, por ejemplo, los Hohenlohen, los Brandemburgo o Baviera. Todas estas casas reales y nobles están fuertemente interconectadas estableciendo una red. En la figura 2.3 se presenta una genealogía parcial de los Habsburgo representada a modo de cadena de descendencia.

**Tabla 2.1.** Resumen de las variables utilizadas en el estudio. En la tabla se muestran los datos completos y agrupados por edades de las mujeres. Para más información, mirar punto 2.4.

	Mínimo	Máximo	Media	Varianza
Toda las familias agrupadas (n=124)				
Variables dependientes				
Fertilidad completada	0	18	4,742	21,965
Periodo Reproductivo Corregido	0	22	6,298	47,642
Variables explicativas				
Coef. Consanguinidad varones	0,000	0,254	0,048	0,004
Coef. Consanguinidad hembras	0,000	0,305	0,041	0,003
Coef. Coascendencia	0,000	0,305	0,051	0,004
Variables nuisance				
Edad del varón en el matrimonio	10	79	28,991	142,853
Edad de la hembra en el matrimonio	11	49	20,750	38,449
Fecha del matrimonio	1452	1771	1618,669	-
Duración del matrimonio	1	30	16,814	82,347
Clase edad de la mujer de 17 a 27 años (n=99)				
Variables dependientes				
Fertilidad completada	0	10	3,182	7,130
Variables explicativas				
Coef. Consanguinidad varones	0,000	0,254	0,048	0,004
Coef. Consanguinidad hembras	0,000	0,305	0,042	0,003
Coef. Coascendencia	0,000	0,305	0,048	0,003
Variables nuisance				
Edad del varón en el matrimonio	16	60	30,43	111,16
Fecha del matrimonio	1452	1771	1620,54	-
Duración del matrimonio	2	10	7,051	7,538
Clase edad de la mujer de 30 a 40 años (n=92)				
Variables dependientes				
Fertilidad completada	0	9	2,434	7,567
Variables explicativas				
Coef. Consanguinidad varones	0,000	0,254	0,043	0,003
Coef. Consanguinidad hembras	0,000	0,156	0,033	0,002
Coef. Coascendencia	0,000	0,254	0,047	0,003
Variables nuisance				
Edad del varón en el matrimonio	24	79	43,835	135,67
Fecha del matrimonio	1452	1771	1623,87	-
Duración del matrimonio	2	10	8,824	5,369

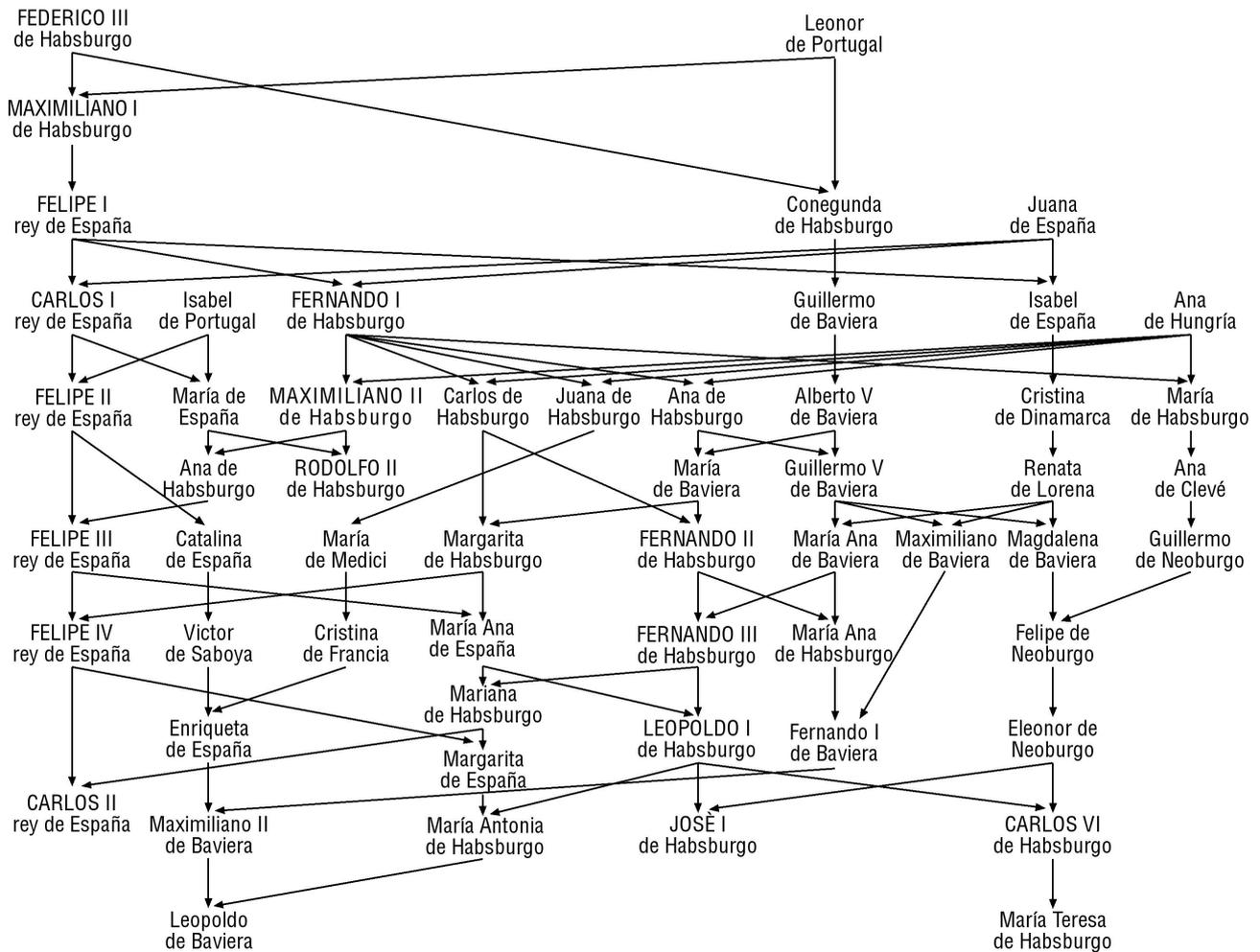


Figura 2.3. Genealogía parcial, representada por cadena de descendencia, de los dos linajes de la dinastía de los Habsburgo: Linaje español y austriaco.

### 2.3.2 Variables nuisance

Para poder obtener conclusiones que tengan sentido y validez biológica es necesario añadir al modelo una serie de variables nuisance que van a permitir controlar y fijar distintos efectos relacionados con la fertilidad pero que no tienen que ver con la consanguinidad de las parejas. A continuación se presenta una pequeña lista con dichas variables.

- Edad del varón en el matrimonio: Presente en todos los modelos analizados.
- Edad de la hembra en el matrimonio: Esta variable se ha utilizado sólo en el análisis de los 124 datos agrupados (ver punto 2.4 Diseño de los Analisis).
- Fecha de Matrimonio: Esta variable permite tener un control temporal. Es un control indirecto de la homogeneidad ambiental a lo largo de los 400 años que abarca el estudio. Se añadió a todos los modelos analizados.

- Duración del Matrimonio: Variable de gran relevancia para la fertilidad de una pareja. Presente en todos los modelos analizados.
- Niños nacidos con anterioridad: Variable utilizada en el análisis del grupo de edad de 30 a 40 años. Indica los niños nacidos en una determinada pareja en el periodo de edad de la mujer anterior a cumplir los 30 años (ver punto 2.4 Diseño de los Análisis).
- Mortalidad infantil sucedida con anterioridad: Variable utilizada en el análisis del grupo de edad de 30 a 40 años. Indica la mortalidad infantil sucedida en el periodo de edad de la mujer anterior a cumplir los 30 años (ver punto 2.4 Diseño de los Análisis).

### 2.4 Diseño de los análisis.

Los datos de fertilidad de las distintas familias de los Habsburgo se van a organizar en distintos conjuntos para poder contestar a preguntas concretas. En el caso de la fertilidad medida como el número de hijos de las parejas – fertilidad completada – los datos se organizaron en 3 conjuntos. El primero, incluye todas las 124 familias estudiadas, sin prescindir de ninguna de ellas. Posteriormente, los análisis se llevaron a cabo dividiendo los datos según dos clases de edad de las mujeres: de 17 a 27 años (n=99) y de 30 a 40 años (n=92). De esta forma, si una mujer se casó a los 16 y se mantuvo casada hasta los 50 sus datos de fertilidad se dividirán en dos clases: niños nacidos en el periodo de edad de la mujer de 17 a 27 años y niños nacidos en el periodo de edad de 30 a 45 años. Esta división de los datos se realiza para poder evaluar el efecto que tiene la senescencia de las mujeres en la depresión consanguínea. En la figura 2.4 se presenta la distribución del número de hijos para las dos clases de datos consideradas según la edad de las mujeres.

En el caso de la otra variable de fertilidad utilizada, periodo reproductivo corregido, se analizaron los datos en su conjunto ya que en el contexto de esta variable no tiene sentido separarlos por grupos de edad.

Además de la relación entre la senescencia de las mujeres y la consanguinidad, también resulta sumamente interesante conocer como fue el comportamiento de la depresión consanguínea en fertilidad a lo largo del tiempo. Para ello, todos los análisis se realizaron dividiendo los datos en tres periodos según la fecha de matrimonio de las parejas. De esta manera primero se analizaron todos los datos, desde el 1400 hasta el 1800 y posteriormente se dividieron en dos periodos: 1400 – 1599 y 1600 – 1800. Esta división temporal es muy útil para analizar si la selección natural pudo eliminar alelos deletéreos de gran efecto, lo que se denomina “purgado”. De ser así, se debería encontrar una depresión consanguínea más acusada en la primera época (1400 – 1599). En caso contrario, estaríamos ante un evento de anti-purgado, donde alelos deletéreos de pequeño efecto – indetectables para la selección natural – se irían acumulando poco a poco.

### 2.5 Métodos estadísticos

El objetivo de la investigación que aquí se presenta es contrastar estadísticamente el efecto de los distintos coeficientes de consanguinidad – varones, mujeres y descendencia – sobre la fertilidad, medida a través del número de hijos y el periodo reproductivo corregido de cada pareja, además, teniendo en cuenta distintas variables nuisance como la edad de los varones y mujeres, así como la fecha y duración

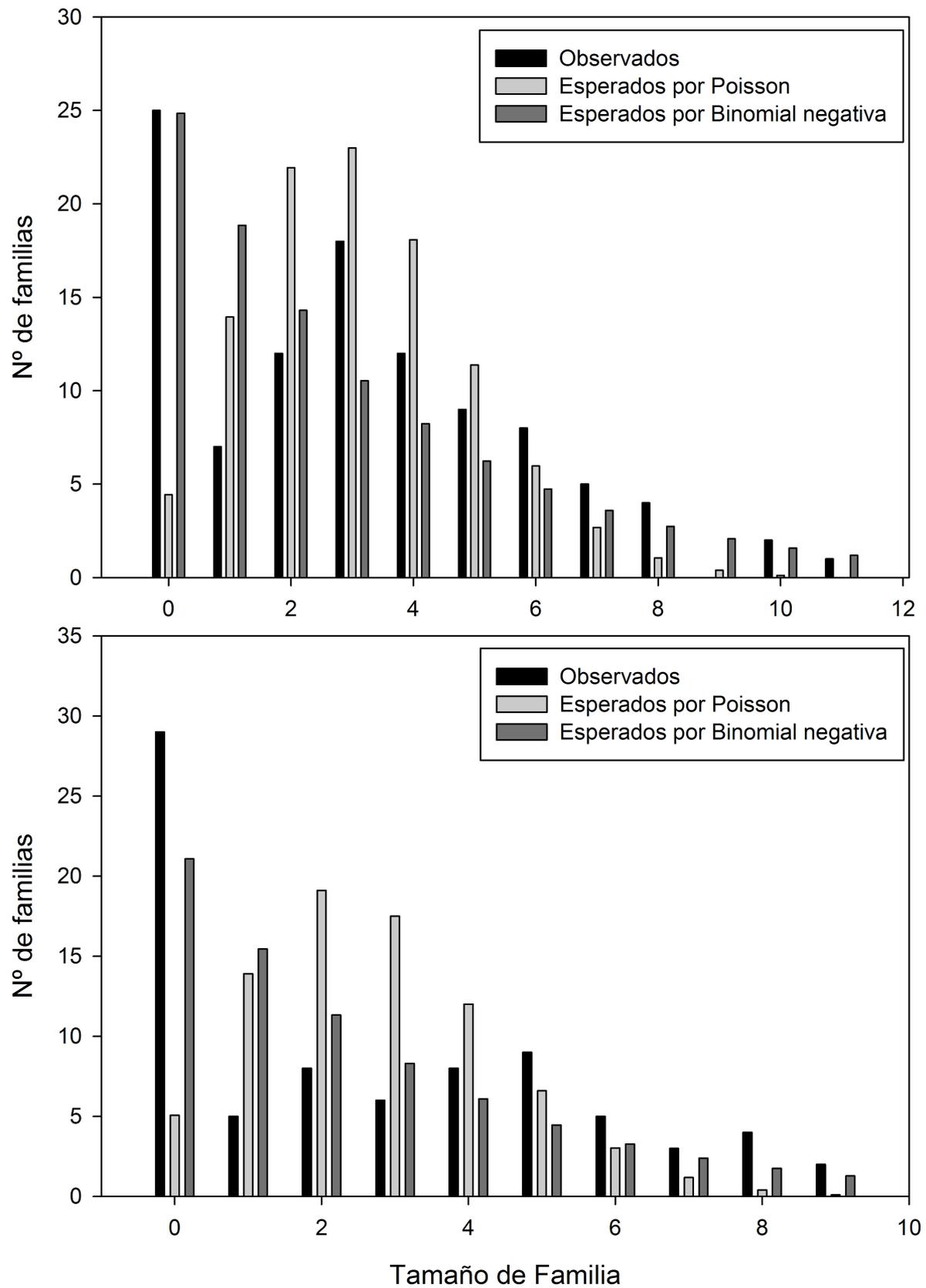


Figura 2.4. Distribución del número de hijos de las parejas Habsburgo, agrupadas según la edad de la madre, junto con la distribución esperada si esta variable de fertilidad siguiese una distribución de poisson o binomial negativa. En la figura superior se presentan los datos del grupo de edad de 17 a 27 años. En la figura inferior se presentan los datos del grupo de edad de 30 a 40 años.

del matrimonio. Este modelo de regresión múltiple no se trata de un análisis exploratorio donde se vaya a proceder a realizar una selección de variables. Este es un análisis confirmatorio, donde biológicamente el modelo se da por cierto y se pretende analizar los posibles efectos de las tres variables de coeficientes de consanguinidad. Para poder resolver esta hipótesis se han utilizado los paquetes `pscl`, `lmtest` y `Hmisc` del entorno estadístico R (R Core Team, 2013).

De esta manera, primero, se van a estimar las funciones de densidad, a través de métodos no paramétricos, de los distintos coeficientes de consanguinidad. Aunque la estimación de estas funciones no es un hito central para los objetivos de esta investigación, su aproximación se hace muy interesante como análisis exploratorio de los datos y para futuras comparaciones con otras dinastías reales. Posteriormente, para estudiar la depresión consanguinidad en fertilidad, se proponen cuatro modelos de regresión: dos modelos GLM (Poisson y binomial negativa) y dos modelos mixturados con eros inflados (poisson cero inflado, ZIP, y binomial negativa cero inflado, ZINB). Tradicionalmente, para analizar la relación entre diversos factores de riesgo y la fertilidad completada o el periodo reproductivo se han utilizado modelos GLM como el Poisson o el binomial negativo si había sobre-dispersión en la variable respuesta (Cavalli-Sforza y Bodmer, 1971; Wood, 1994). Por otra parte, en la muestra a analizar en este caso, existen un gran número de ceros en las variables respuesta, bastantes más de los esperados por las distribuciones de Poisson o Binomiales Negativas (figuras 2.1, 2.2, 2.4). A pesar de que este exceso de ceros se ha observado en la distribución marginal de las variables respuestas, parece lógico aplicar modelos mixtos por ceros inflados.

### 2.5.1 Estimación de la densidad del coeficiente de consanguinidad

Debido a que se desconoce cómo es la distribución que siguen los coeficientes de consanguinidad de los varones, las mujeres y las parejas en las dinastías reales europeas de la Edad Moderna, se van a estimar sus funciones de densidad mediante dos estimadores no paramétricos. Estos métodos son de gran ayuda cuando no se puede asumir que la variable bajo estudio pertenece a una determinada familia paramétrica. De hecho, los estimadores paramétricos son los más eficientes cuando los datos se adaptan correctamente a una distribución paramétrica dada. Sin embargo, si no proceden de dicha distribución estos estimadores son inconsistentes. Los estimadores no paramétricos son generalmente consistentes, no obstante, suelen ser menos eficientes que los estimadores paramétricos cuando el tamaño de muestra es pequeño.

Para realizar un análisis exploratorio de los datos se va a utilizar el histograma que es el estimador no paramétrico clásico para la función de densidad. Sin embargo, uno de los principales problemas del histograma como estimador de la función de densidad es que este no es continuo. Este problema puede solventarse con el estimador Kernel o de tipo núcleo. Este se define como:

$$\hat{f}_{n,K}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (2.2)$$

donde  $K$  es la función Kernel que debe cumplir las condiciones de ser una función de densidad, ser simétrica con respecto al origen y con momento de orden cuatro finito.  $h$  es el denominado parámetro ventana, es el que regula el suavizado de la curva de densidad. La elección de la función Kernel es poco

irrelevante en la estimación de la función de densidad (Wand y Jones, 1995. p. 28-31; Simonof, 1996. p. 41-44). En los datos aquí analizados hemos tomado una función Kernel gaussiana.

Para el caso particular de la distribución de los coeficientes de consanguinidad es necesario introducir una transformación en la variable, ya que el estimador de tipo núcleo tiene un soporte no acotado, es decir, está definido en toda la recta real, mientras que los coeficientes de consanguinidad toman valores entre 0 y 1. Para ello se transformaron logarítmicamente los coeficientes de consanguinidad de los varones las mujeres y la descendencia. Si  $f$  denota la densidad de  $X$  y  $g$  denota la densidad de la variable transformada  $T(X)$  tenemos que:

$$f(\mathbf{x}) = g(T(\mathbf{x}))T'(\mathbf{x}) \quad (2.3)$$

Por lo que  $g$  puede ser estimada por el método Kernel estándar

$$\hat{f}_{n,K}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n K_h(T(\mathbf{x}) - T(X_i))T'(\mathbf{x}) \quad (2.4)$$

Para la selección del parámetro ventana ( $h$ ) se utilizó el método “plug – in” desarrollado por Sheather y Jones (1991) (Sheather, 2004). Este selector busca el parámetro ventana ( $h$ ) que minimice la aproximación asintótica del error cuadrado medio integrado (AMISE).

### 2.5.2 Modelos Generales Linearizados (GLM)

Los modelos lineales generalizados fueron formulados por Robert W. M. Wedderburn y John A. Nelder, dos estadísticos británicos, en la segunda mitad del siglo pasado (Nelder y Wedderburn, 1972) como manera de unificar varios modelos estadísticos incluyendo la regresión lineal, regresión logística y la regresión de Poisson bajo un solo marco teórico. Los modelos lineales de regresión se basan en ciertos supuestos como la relación lineal entre la variable respuesta y las variables independientes, varianza constante y errores con distribución normal, sin embargo, en muchas ocasiones, una o varias de estos supuestos no se cumplen. Estos problemas pueden solventarse con transformaciones matemáticas de las variables, no obstante, estas soluciones tienen limitaciones y no siempre consiguen corregir la falta de normalidad, heterocedasticidad o linealidad. Una alternativa a la transformación de las variables respuesta es el uso de los modelos lineales generalizados (GLM en sus siglas en inglés: generalized linear models). Estos modelos son extensiones flexibles de los modelos lineales que permiten utilizar distribuciones no normales de los errores (Poisson, binomiales, etc) y varianzas no constantes. Otra razón por la que los modelos lineales generalizados son necesarios es que la suposición de relación lineal entre la variable respuesta y variables independientes no siempre se cumple. Este problema tiene solución en los GLM ya que estos no modelizan la esperanza de  $Y$ , sino una función de la esperanza de  $Y$   $g()$  que se conoce como función de enlace, de la familia exponencial, que se encarga de linearizar la relación entre las variables. De esta forma definiríamos el modelo como:

$$E(\mathbf{Y}) = \boldsymbol{\mu} = \mathbf{g}^{-1}(\mathbf{X}\boldsymbol{\beta}) \quad (2.5)$$

Donde  $E(Y)$  es el valor esperado de  $Y$ ,  $X\beta$  es el predictor lineal, una combinación lineal de parámetros desconocidos  $\beta$  multiplicados por las variables explicativas y  $g$  es la función de enlace de la familia exponencial. Los parámetros  $\beta$  son generalmente estimados por máxima verosimilitud. De esta forma, para una variable respuesta conteo, por ejemplo, la distribución de los errores seguirá una Poisson o, en el caso de haber sobre-dispersión, una binomial negativa, en los dos casos la función enlace sería logarítmica.

Para analizar el efecto que tiene sobre las variables de fertilidad de los Habsburgo los distintos coeficientes de consanguinidad es necesario utilizar modelos lineales generalizados, esto es debido a que las dos variables respuesta que se están considerando son recuentos con soporte positivo  $(0, +\infty)$ . Es por ello por lo que se han utilizado modelos de regresión por Poisson y binomial negativa; este último modelo es utilizado en muchas áreas de la investigación biológica para lidiar con la sobre-dispersión (White y Bennets, 1996; Lloyd, 2007). En el caso que aquí se presenta, tanto la varianza de la fertilidad completada como del periodo reproductivo corregido son más del doble de sus respectivas medias, por lo que estamos ante un caso de sobre-dispersión (Zurr et al., 2009).

De esta manera se definen los modelos a utilizar

### 2.5.2.1 Poisson

La regresión de Poisson tiene por función de enlace la logarítmica, por lo que se define:  $X\beta = \ln(\mu)$ . Por lo que, si se define  $\lambda(x)=E(Y/X=x)$  la función de regresión del modelo de Poisson quedaría expresado como:  $\lambda(x,\beta)=e^{x\beta}$ .

Para el análisis del conjunto de todos los datos (124 familias) se define la muestra  $(X_1, Y_1), \dots, (X_n, Y_n)$

Donde:

$Y_i$  = número de hijos

$X_i = (X_{i1}, \dots, X_{ip})$

$X_{i1}$ : coeficiente de consanguinidad de los padres ( $F_S$ )

$X_{i2}$ : coeficiente de consanguinidad de las madres ( $F_D$ )

$X_{i3}$ : coeficiente de coascendencia de las parejas ( $T_i$ )

$X_{i4}$ : año de matrimonio (Mat)

$X_{i5}$ : edad del varón en el matrimonio (AgeS)

$X_{i6}$ : edad de la hembra en el matrimonio (AgeD)

$X_{i7}$ : duración del matrimonio. (DM)

$Y_i \in \text{Poisson}(\lambda(X_i, \beta))$

$$\log \lambda(X_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{AgeD} + \beta_7 \text{DM} \quad (2.6)$$

En el caso del análisis de los datos agrupados según dos periodos de edades de las progenitoras (17 a 27 años y 30 a 40 años) los modelos son ligeramente distintos a los anteriores.

- Para el conjunto de datos correspondiente al grupo de mujeres de 17 a 27 años ( $n=99$ )

$Y_i$  = número de hijos

$X_i = (X_{i1}, \dots, X_{ip})$

$X_{i1}$ : coeficiente de consanguinidad de los padres ( $F_S$ )

$X_{i2}$ : coeficiente de consanguinidad de las madres ( $F_D$ )

$X_{i3}$ : coeficiente de coascendencia de las parejas ( $T_i$ )

$X_{i4}$ : año de matrimonio (Mat)

$X_{i5}$ : edad del varón en el matrimonio (AgeS)

$X_{i6}$ : duración del matrimonio. (DM)

$Y_i \in \text{Poisson}(\lambda(X_i, \beta))$

$$\log \lambda(X_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{DM} \quad (2.7)$$

- Para el conjunto de datos correspondiente al grupo de mujeres de 30 a 40 años ( $n=92$ )

$Y_i$  = número de hijos

$X_i = (X_{i1}, \dots, X_{ip})$

$X_{i1}$ : coeficiente de consanguinidad de los padres ( $F_S$ )

$X_{i2}$ : coeficiente de consanguinidad de las madres ( $F_D$ )

$X_{i3}$ : coeficiente de coascendencia de las parejas ( $T_i$ )

$X_{i4}$ : año de matrimonio (Mat)

$X_{i5}$ : edad del varón en el matrimonio (AgeS)

$X_{i6}$ : duración del matrimonio. (DM)

$X_{i7}$ : niños nacidos con anterioridad. (N)

$X_{i8}$ : mortalidad infantil en la etapa anterior. (M)

$Y_i \in \text{Poisson}(\lambda(X_i, \beta))$

$$\log \lambda(X_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{DM} + \beta_7 N + \beta_8 M \quad (2.8)$$

En el caso del periodo reproductivo corregido (PRC), debido a que el análisis se realiza de todos los datos en su conjunto tenemos sólo un modelo:

$$Y_i = \text{PRC}$$

$$X_i = (X_{i1}, \dots, X_{ip})$$

$X_{i1}$ : coeficiente de consanguinidad de los padres ( $F_S$ )

$X_{i2}$ : coeficiente de consanguinidad de las madres ( $F_D$ )

$X_{i3}$ : coeficiente de coascendencia de las parejas ( $T_i$ )

$X_{i4}$ : año de matrimonio (Mat)

$X_{i5}$ : edad del varón en el matrimonio (AgeS)

$X_{i6}$ : edad de la hembra en el matrimonio (AgeD)  $X_{i7}$ : duración del matrimonio. (DM)

$$Y_i \in \text{Poisson}(\lambda(X_i, \beta))$$

$$\log \lambda(X_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{AgeD} + \beta_7 \text{DM} \quad (2.9)$$

### 2.5.2.1 Binomial Negativa

El modelo de la distribución binomial negativa es adecuado para representar situaciones de sobre-dispersión, donde la varianza es mayor que la media; por ello se va a parametrizar a partir de su media ( $\lambda$ ) y de su sobre-dispersión ( $k$ ). Entonces se tiene que  $E(y) = \lambda$  y que la  $\text{Var}(y) = \lambda + \lambda^2/k$ . De esto se deriva que cuanto mayor sea  $k$  menor será la sobre-dispersión, si  $k \rightarrow \infty$  la distribución binomial converge a una Poisson con varianza y media iguales. El modelo de regresión binomial negativo se puede expresar como  $\lambda(x, \beta) = e^{x'\beta}$ . Esto es, la media depende de la variable explicativa a través de una función log-lineal, mientras que el parámetro de sobre-dispersión  $k$  es constante, por lo que no depende de variables explicativas. La distribución binomial negativa solía verse como una combinación de dos distribuciones: se asume que las respuestas (las  $Y$ ) se distribuyen bajo una Poisson, y que la media  $\mu$  se distribuye como una gamma (Hilbe, 2007). Según esto, el modelo de regresión de Poisson esta encajado en el modelo de binomial negativa, por lo que ambos modelos se pueden comparar mediante un test de razón de verosimilitud.

En el caso de la regresión binomial negativa las variables son las mismas que en los casos anteriormente vistos.

- Para el conjunto todos los datos ( $n=124$ ).

$$Y_i \in \text{BN}(\lambda(X_i, \beta), k)$$

$$\log \lambda(X_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{AgeD} + \beta_7 \text{DM} \quad (2.10)$$

- Para el conjunto de datos correspondiente a mujeres de 17 a 27 años ( $n=99$ )

$$Y_i \in \text{BN}(\lambda(X_i, \beta), k)$$

$$\log \lambda(X_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{DM} \quad (2.11)$$

- Para el conjunto de datos correspondiente a mujeres de 30 a 40 años ( $n=92$ )

$$Y_i \in \text{BN}(\lambda(X_i, \beta), k)$$

$$\log \lambda(X_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{DM} + \beta_7 N + \beta_8 M \quad (2.12)$$

En el caso del periodo reproductivo corregido (PRC), el modelo es:

$$Y_i \in \text{BN}(\lambda(X_i, \beta), k)$$

$$\log \lambda(X_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{AgeD} + \beta_7 \text{DM} \quad (2.13)$$

### 2.5.3 Modelos Mixturados. Modelos con ceros inflados

Las variables de recuento son a veces caprichosas con la frecuencia de aparición de ceros, de esta manera, se pueden encontrar variables en las que no es posible obtener un recuento de ceros (ceros truncados), o variables donde la frecuencia de ceros es mucho mayor a la esperada por el modelo (ceros inflados o apartados). Para tratar con estas variables respuesta, es necesario realizar modificaciones en los modelos lineales generales. En el caso de la usencia de ceros se deben realizar modificaciones dirigidas a que los modelos no otorguen una probabilidad positiva al cero. Para ello se modifica el valor de  $\lambda$  de la Poisson de modo que ya no coincide con su media, y se realizan modificaciones en la binomial negativa de modo que  $P(Y=y) = P(Z=y/Z>0)$  si  $y \in \{1, 2, \dots\}$ .

Otro problema distinto se presenta cuando la variable respuesta presenta más ceros de los esperados por los distintos modelos (Poisson y binomial negativa). En estos casos es necesario utilizar los llamados modelos mixturados que tienen la habilidad de poder modelizar dos poblaciones a la vez. Es por ello que es de vital importancia determinar correctamente la procedencia de la masa de ceros, se pueden dar dos posibilidades. La primera, denominada ceros apartados, se caracteriza por que la masa de ceros está originada por un proceso subyacente distinto al que genera los conteos. El modelo de cero apartado consiste en separar los ceros del resto de observaciones, de modo que  $\pi$  sería la probabilidad de cero, mientras que para las demás observaciones se emplearía una distribución truncada en el cero. La segunda posibilidad, denominada ceros inflados, se caracteriza por que la masa de ceros proviene de dos procesos distintos: una parte de los ceros son ceros “estructurales” o “falsos” ceros y la otra parte provienen de la distribución de Poisson o binomial negativa de manera natural. Por lo tanto, se determina la probabilidad de observar un cero estructural con  $\pi$  y con probabilidad complementaria  $(1-\pi)$  obtener una observación de la distribución de Poisson o binomial negativa. Las medias de estos procesos también se verán afectadas como consecuencia del cero inflado. De esta forma, la media de la distribución inflada verifica:  $E(Y)=E(Z) (1-\pi)$ . Si  $Z$  pertenece a una Poisson inflada entonces:  $E(Y)=\lambda(1-\pi)$ , mientras que si

pertenece a una binomial negativa inflada entonces:  $E(Y)=\mu(1-\pi)$ . En un modelo de regresión, tanto  $\mu$  como  $\pi$  se pueden expresar en función de las variables explicativas. En el caso de  $\mu$  se utilizaría un modelo log-lineal, mientras que en el caso de  $\pi$  se utilizaría un moldeo de regresión logístico.

Pese a que tanto el número total de hijos como el periodo reproductivo corregido son variables recuento, las distribuciones marginales de estas variables se pueden encontrar muchos más ceros que los esperados por los modelos poisson y binomial negativa. Ignorar esta gran masa de ceros en la distribución marginal de la variable respuesta puede traer dos consecuencias; primero, la estima de los distintos parámetros de los modelos puede ser sesgada, y segundo, el número excesivo de ceros puede causar sobre-dispersión. En el caso que aquí se está analizando, la procedencia de los ceros es mixta; una familia puede tener cero hijos por múltiples causas tanto genéticas como no genéticas, es más, dentro de las causas genéticas estas pueden estar relacionadas o no con la consanguinidad. Lo mismo sucede en el periodo reproductivo de una familia. De esta manera, la masa de ceros puede provenir de dos fuentes: de una distribución de poisson o binomial negativa como el resto de datos, y además, de otros procesos independientes. Es por ello por lo que la utilización de modelos mixtos por ceros inflados como ZIP o ZIMB (zero inflated poisson y zero inflated negative binomial respectivamente) parece ser la más adecuada. De esta manera, el modelo poisson con ceros inflados (ZIP) puede soportar la posible sobre-dispersión generada por la acumulación de ceros, mientras que el modelo binomial negativa con ceros inflados (ZINB) puede soportar la sobre-dispersión creada por la masa excesiva de ceros y además la propia generada por los datos de recuento (Zeileis et al., 2008; Zuur et al., 2009). Como en el caso de los modelos GLM más simples, el modelo ZIP esta encajado en el modelo ZINB, por lo que se puede establecer una comparación mediante un test de razón de verosimilitud.

Desde un punto de vista biológico, la aplicación de modelos mixturados con ceros inflados en el estudio de los efectos de la consanguinidad sobre la fertilidad tiene la ventaja de que se puede resolver dos cuestiones a la vez: la fertilidad, medida como el número de hijos de las parejas o el periodo reproductivo – que se modeliza en la parte del modelo de regresión por contaje – y la infertilidad, representada por esa masa de ceros alternativa y modelizada en los ceros inflados. De esta manera, los modelos de ceros inflados permiten conocer al mismo tiempo si la consanguinidad tuvo un efecto significativo sobre el número de hijos o periodo reproductivo de las parejas fértiles y además sobre la condición de las parejas infértiles. Este trabajo es el primero del que se tiene constancia que utiliza los modelos mixturados en el estudio de los efectos de la consanguinidad en el hombre.

### 2.5.3.1 Poisson cero inflado

Si  $Z \in \text{Poisson}(\lambda)$ , la observación  $Y$  con ceros inflados tendría distribución:

si  $y \in \{1,2,3,\dots\}$

$$P(Y = 0) = \pi(x_i, \gamma) + (1 - \pi(x_i, \gamma))e^{-\lambda(x_i, \beta)} \quad (2.14)$$

si  $y \in \{1,2,3,\dots\}$

$$P(Y_i = y) = (1 - \pi(x_i, \gamma)) e^{-\lambda(x_i, \beta)} \frac{\lambda(x_i, \beta)^y}{y!} \quad (2.15)$$

Para el análisis de todos los datos se define:

$$\log \lambda(x_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{AgeD} + \beta_7 \text{DM} \quad (2.16)$$

$$\text{logit}(\pi(x_i, \gamma)) = \gamma_0 + \gamma_1 F_S + \gamma_2 F_D + \gamma_3 T_i + \gamma_4 \text{Mat} + \gamma_5 \text{AgeS} + \gamma_6 \text{AgeD} + \gamma_7 \text{DM} \quad (2.17)$$

Donde:

$$\text{logit}(\pi(x_i, \gamma)) = \log \frac{\pi}{1 - \pi} \quad (2.18)$$

Para el conjunto de datos correspondiente a mujeres de 17 a 27 años (n=99) se define:

$$\log \lambda(x_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{DM} \quad (2.19)$$

$$\text{logit}(\pi(x_i, \gamma)) = \gamma_0 + \gamma_1 F_S + \gamma_2 F_D + \gamma_3 T_i + \gamma_4 \text{Mat} + \gamma_5 \text{AgeS} + \gamma_7 \text{DM} \quad (2.20)$$

Para el conjunto de datos correspondiente a mujeres de 30 a 40 años (n=92) se define:

$$\log \lambda(x_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{DM} + \beta_7 \text{Na} + \beta_8 \text{Ma} \quad (2.21)$$

$$\text{logit}(\pi(x_i, \gamma)) = \gamma_0 + \gamma_1 F_S + \gamma_2 F_D + \gamma_3 T_i + \gamma_4 \text{Mat} + \gamma_5 \text{AgeS} + \gamma_6 \text{DM} + \gamma_7 \text{N} + \gamma_8 \text{M} \quad (2.22)$$

Para los análisis del periodo reproductivo corregido (PRC), se define:

$$\log \lambda(x_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{AgeD} + \beta_7 \text{DM} \quad (2.23)$$

$$\text{logit}(\pi(x_i, \gamma)) = \gamma_0 + \gamma_1 F_S + \gamma_2 F_D + \gamma_3 T_i + \gamma_4 \text{Mat} + \gamma_5 \text{AgeS} + \gamma_6 \text{AgeD} + \gamma_7 \text{DM} \quad (2.24)$$

### 2.5.3.2 Binomial Negativa cero inflado

Si  $Z \in$  Binomial negativa  $(\lambda, k)$ , la observación  $Y$  con ceros inflados tendría distribuciones donde la  $P(Y=0)$  y  $P(Y=y)$  serían iguales a las expresiones (2.14) y (2.15), respectivamente, sustituyendo la expresión de la Poisson por la expresión de la binomial negativa donde:

si  $y \in \{1, 2, 3, \dots\}$

$$P(Y = y) = \frac{\Gamma(k + y)}{\Gamma(k)y!} \frac{\lambda^y k^k}{(\lambda + k)^{k+y}} \quad (2.25)$$

Para el análisis de todos los datos se define:

$$\log \lambda(x_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{AgeD} + \beta_7 \text{DM} \quad (2.26)$$

$$\text{logit}(\pi(x_i, \gamma)) = \gamma_0 + \gamma_1 F_S + \gamma_2 F_D + \gamma_3 T_i + \gamma_4 \text{Mat} + \gamma_5 \text{AgeS} + \gamma_6 \text{AgeD} + \gamma_7 \text{DM} \quad (2.27)$$

Para el conjunto de datos correspondiente a mujeres de 17 a 27 años ( $n=99$ ) se define:

$$\log \lambda(x_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{DM} \quad (2.28)$$

$$\text{logit}(\pi(x_i, \gamma)) = \gamma_0 + \gamma_1 F_S + \gamma_2 F_D + \gamma_3 T_i + \gamma_4 \text{Mat} + \gamma_5 \text{AgeS} + \gamma_7 \text{DM} \quad (2.29)$$

Para el conjunto de datos correspondiente a mujeres de 30 a 40 años ( $n=92$ ) se define:

$$\log \lambda(x_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{DM} + \beta_7 N + \beta_8 M \quad (2.30)$$

$$\text{logit}(\pi(x_i, \gamma)) = \gamma_0 + \gamma_1 F_S + \gamma_2 F_D + \gamma_3 T_i + \gamma_4 \text{Mat} + \gamma_5 \text{AgeS} + \gamma_6 \text{DM} + \gamma_7 N + \gamma_8 M \quad (2.31)$$

Para los análisis del periodo reproductivo corregido (PRC), se define:

$$\log \lambda(x_i, \beta) = \beta_0 + \beta_1 F_S + \beta_2 F_D + \beta_3 T_i + \beta_4 \text{Mat} + \beta_5 \text{AgeS} + \beta_6 \text{AgeD} + \beta_7 \text{DM} \quad (2.32)$$

$$\text{logit}(\pi(x_i, \gamma)) = \gamma_0 + \gamma_1 F_S + \gamma_2 F_D + \gamma_3 T_i + \gamma_4 \text{Mat} + \gamma_5 \text{AgeS} + \gamma_6 \text{AgeD} + \gamma_7 \text{DM} \quad (2.33)$$

# ● Resultados



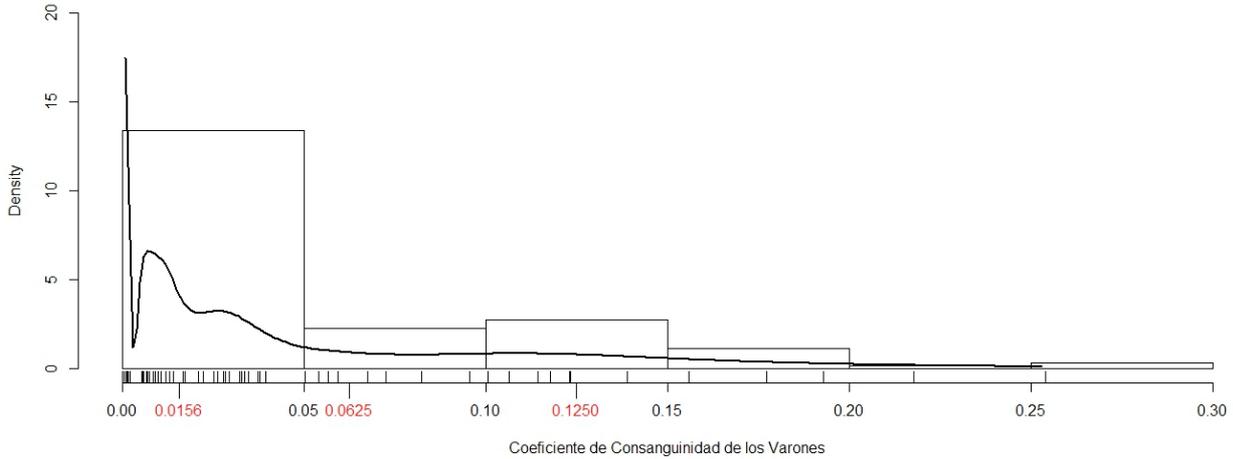
### 3. RESULTADOS

En el presente trabajo se investigaron los efectos de la consanguinidad sobre la fertilidad de 124 parejas de la dinastía real europea de los Habsburgo, que vivieron entre 1400 y 1800. Para ello primero se realizó un análisis exploratorio de los coeficientes de consanguinidad de los varones, mujeres y las parejas mediante la estimación no paramétrica de la densidad de dichos coeficientes. Posteriormente, para estudiar el efecto del alta consanguinidad de los Habsburgo en su fertilidad se han utilizado dos medidas de la fertilidad: el número de hijos por pareja – fertilidad completada – y el periodo reproductivo corregido (PRC). Como variables explicativas de interés se utilizaron los coeficientes de consanguinidad de los varones, las mujeres y sus descendencias ( $F_S$ ,  $F_D$  y  $T_i$  respectivamente). Además se introdujeron una serie de variables nuisance que permiten descontar el efecto de condicionantes biológicos y ambientales que poco tienen que ver con la consanguinidad pero mucho con la fertilidad de las parejas. Como se ha afirmado en los materiales y métodos los distintos modelos de regresión utilizados no son modelos exploratorios, sino que son modelos confirmatorios, por lo que no se aplicarán métodos de selección de variables ya que todas ellas tienen un sentido biológico en el modelo.

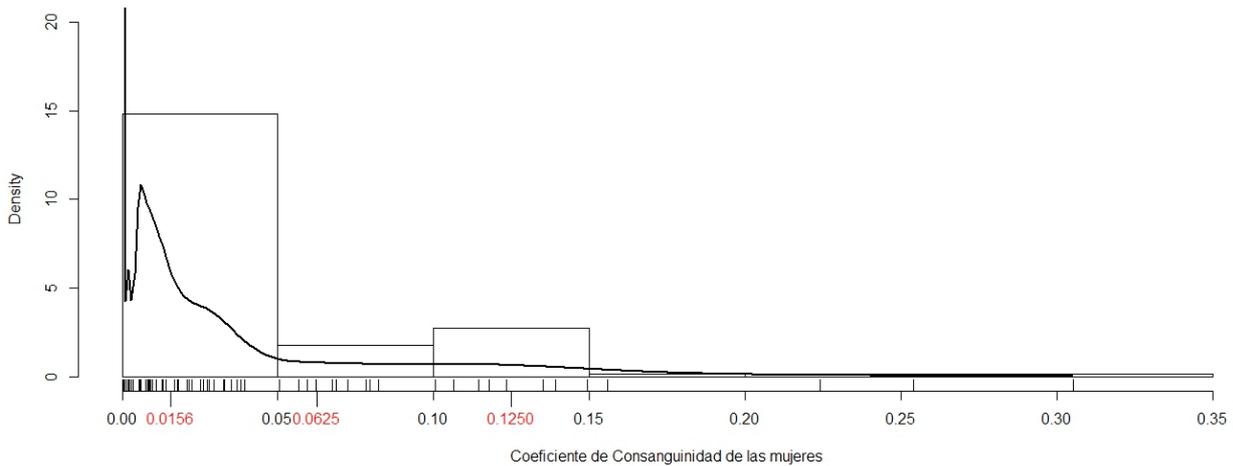
#### 3.1. Estimación no paramétrica de la densidad de los coeficientes de consanguinidad

Las familias Habsburgo tuvieron coeficientes de consanguinidad que abarcan un rango de 0 a 0.305, es decir, en este periodo y en esta dinastía, hubo desde enlaces no consanguíneos hasta uniones entre parientes con una relación más cercana que un grado incestuoso: hermano – hermana, padre – hija, madre – hijo. La consanguinidad media de los varones de la dinastía fue de 0.048, la de las mujeres 0.041 y la de la descendencia 0.051; estos valores son un poco más bajos de los que corresponden a un enlace entre primos ( $F = 0.0625$ ). En las figuras 3.1, 3.2 y 3.3 se representan respectivamente las estimaciones no paramétricas de las densidades de los coeficientes de consanguinidad de los varones, las mujeres y la descendencia (que es el coeficiente de consanguinidad de las parejas, la coascendencia). En estas figuras se puede observar como los tres coeficientes tienen densidades muy parecidas; las tres estimaciones son asimétricas a la derecha y se componen de distintas sub-poblaciones. La sub-población mayoritaria en los tres coeficientes la componen individuos con una consanguinidad de 0 o prácticamente 0. En este grupo se encuentran individuos con coeficientes de consanguinidad de milésimas; esta consanguinidad procede de enlaces endogámicos sucedidos en antepasados remotos. La siguiente sub-población se encuentra justo antes de la marca del enlace entre primos segundos ( $F=0.0125$ ). Estos individuos son descendientes de antepasados consanguíneos lejanos como primos terceros o primos cuartos, así como antepasados comunes remotos cuya influencia se hace notar a partir del cuarto decimal del coeficiente de consanguinidad. Entre las marcas de los enlaces entre primos segundo y primos primeros ( $F=0.0625$ ) se encuentra otra sub-población. Esta la componen individuos cuyos progenitores son primos segundos pero que además comparten antepasados lejanos comunes cuya incidencia se produce a partir del tercer decimal de los coeficientes de consanguinidad. A partir de la marca del enlace de primos hermanos los datos están mucho más dispersos a lo largo del eje de las x, esto se explica porque cuanto más cercanos sean los enlaces consanguíneos, en la genealogía de un individuo, más peso tendrán en el coeficiente de consanguinidad de dicho individuo, ya que la incidencia de los eventos más remotos se notarán a partir del 3 o 4 decimal.

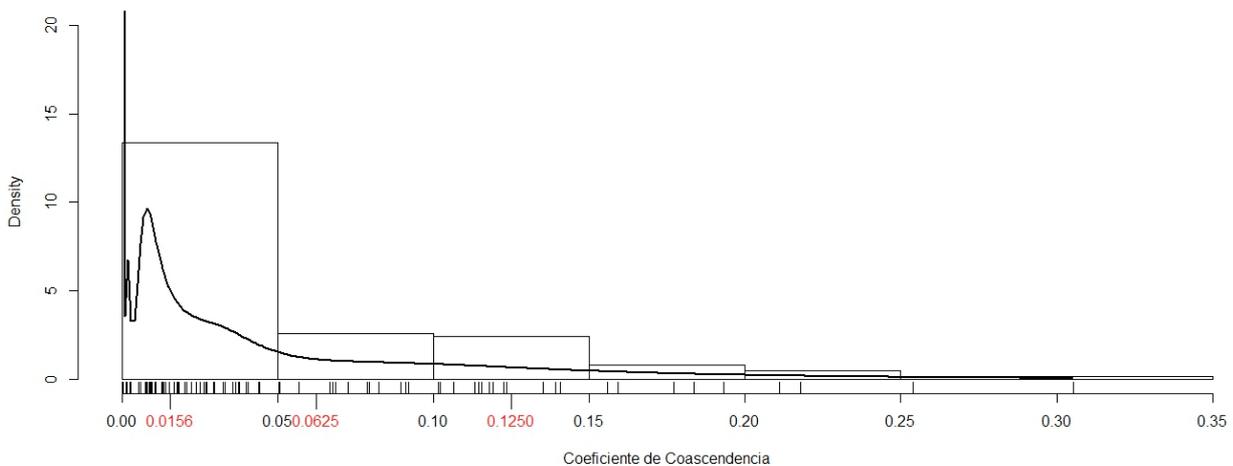
## Resultados



**Figura 3.1.** Estimación no paramétrica de la densidad de los coeficientes de consanguinidad de los varones. Para la selección del parámetro ventana se utiliza el método “plug-in”. En rojo se marcan los coeficientes de consanguinidad de un enlace entre primos segundos ( $F=0.0156$ ), primos hermanos ( $F=0.0625$ ) y tío-sobrino ( $F=0.125$ ).



**Figura 3.2.** Estimación no paramétrica de la densidad de los coeficientes de consanguinidad de las mujeres. Para la selección del parámetro ventana se utiliza el método “plug-in”. En rojo se marcan los coeficientes de consanguinidad de un enlace entre primos segundos ( $F=0.0156$ ), primos hermanos ( $F=0.0625$ ) y tío-sobrino ( $F=0.125$ ).



**Figura 3.3.** Estimación no paramétrica de la densidad de los coeficientes de coascendencia de las parejas. Para la selección del parámetro ventana se utiliza el método “plug-in”. En rojo se marcan los coeficientes de consanguinidad de un enlace entre primos segundos ( $F=0.0156$ ), primos hermanos ( $F=0.0625$ ) y tío-sobrino ( $F=0.125$ ).

Analizando las figuras 3.1, 3.2 y 3.3 se hace patente que el cálculo del coeficiente de consanguinidad, como variable física, definida como la probabilidad de que un individuo herede dos alelos idénticos por descendencia, depende de la información genealógica del individuo en cuestión. De esta manera, cuanto menos información genealógica se disponga el coeficiente de consanguinidad se comportará de una manera más discreta, ya que sólo podrá tomar ciertos valores, a su vez, cuanto más información genealógica se disponga el cálculo de la consanguinidad será más precisa y continua. El coeficiente de consanguinidad también depende en gran medida de cuando se producen los enlaces consanguíneos; los enlaces entre parientes cercanos (primos primeros en adelante) tienen mucho peso en el coeficiente de consanguinidad, mientras que los enlaces consanguíneos lejanos o remotos sólo tienen efecto en los últimos decimales. Este efecto desproporcionado de la consanguinidad cercana fue descrito por distintos autores que afirmaron que una genealogía con una profundidad de 5 generaciones debería capturar el 90% del coeficiente de consanguinidad (Ballou, 1997; Balloux et al., 2004). Sin embargo, Alvarez et al., 2009 y Ceballos y Alvarez, 2013, determinaron que en genealogías complejas el peso en el coeficiente de consanguinidad de un individuo a partir de la 5 generación de profundidad puede ser del 50%. Por todo ello, para obtener una medida fiable y lo más exacta posible del coeficiente de consanguinidad, es necesario disponer de una genealogía con una profundidad adecuada. En el caso que se presenta en este trabajo, todos los individuos analizados disponen de una genealogía con una profundidad de más de 10 generaciones.

La comparación de la estimación no paramétrica de la densidad de los coeficientes de consanguinidad entre distintas dinastías reales de la época – como los Valois, los Borbones, los Avise, etc – va a ser de gran interés ya que estas densidades son un reflejo de la política matrimonial que las dinastías llevaron a cabo a lo largo de la Edad Moderna europea. Además, el rango tan amplio que presentan los coeficientes de consanguinidad de la dinastía Habsburgo permite aproximarse a los efectos de la consanguinidad en la fertilidad humana como no se ha podido hacer hasta ahora, ya que, actualmente, en el género humano es complicado encontrar una muestra lo suficientemente grande y homogénea que llegue hasta grados de endogamia tan altos.

### 3.2. Efectos de la consanguinidad sobre la fertilidad completada

Para analizar los efectos de la consanguinidad sobre la fertilidad completada – número de hijos – de las parejas Habsburgo, los datos se analizaron primero en su conjunto y después clasificándolos según la edad de las mujeres, de 17 a 27 años y de 30 a 40 años. De esta manera se podrá analizar la relación entre la senescencia y los efectos en fertilidad de la consanguinidad. En el primer análisis se analizaron 124 parejas mientras que los grupos de mujeres de 17 a 27 y 30 a 40 años se componen de 99 y 92 parejas cada uno.

#### 3.2.1. Análisis de las 124 familias en su conjunto.

En la tabla 3.1 se presentan los resultados obtenidos para todas las familias analizadas en todo el periodo de estudio (1400 – 1800). Se puede apreciar como el modelo que mejor se ajusta a los datos – según el AIC (607.24) – es el modelo ZINB, siendo el parámetro  $k$  significativo ( $P < 0.000$ ). Así mismo, el test de razón de verosimilitud, entre la ZIP y la ZINB resultó ser también altamente significativo ( $P < 0.000$ ) con lo que se confirma que el modelo que mejor se adapta a los datos es el ZINB. La tabla 3.1 muestra un efecto significativo, de signo negativo, de la coascendencia de las parejas sobre su propia infertilidad, es

decir, cuanto más consanguínea sea la pareja – o lo que es lo mismo, su descendencia – menor probabilidad de que dicha pareja sea infértil (coef = -15,516,  $P=0,032$ ). Este efecto se podría explicar por un proceso de compensación reproductora; se ha demostrado que en esta dinastía las parejas más consanguíneas sufrieron mayor cantidad de muertes infantiles, por lo que dichas familias pudieron mostrar más ímpetu en tener descendencia (Ceballos y Álvarez, 2013). De las variables nuisance utilizadas, sólo la duración del matrimonio tuvo un efecto significativo, tanto en el número de hijos que tuvieron las parejas como en la cantidad de parejas infértiles.

**Tabla 3.1.** Análisis de regresión para todo el set de datos ( $n=124$ ) de la fertilidad completada – número de hijos – en función de los coeficientes de consanguinidad de los progenitores y coascendencia ( $F_S$ ,  $F_D$ ,  $T_i$ , respectivamente), la fecha del matrimonio ( $Mat$ ), la edad media de los varones ( $Age_S$ ), la edad media de las hembras ( $Age_D$ ) y la duración del matrimonio ( $DM$ ). Se utilizaron modelos GLM por Poisson (expresión 2.6), Binomial negativa (expresión 2.10), además de modelos mixturados ZIP (expresiones 2.16, 2.17) y ZINB y (expresiones 2.26, 2.27).  $K$  es el parámetro de dispersión de la Binomial negativa y el AIC es el criterio de información de Akaike.

	ZIP		ZINB		Poisson		Binom. Neg.	
	Esti.	P-val	Esti.	P-val	Esti.	P-val	Esti.	P-val
Modelo por contaje								
$F_S$	-0,284	0,779	-0,069	0,168	-2,063	0,019	-1,792	0,323
$F_D$	0,677	0,563	1,153	0,475	-0,374	0,732	0,556	0,800
$T_i$	-1,095	0,286	-1,444	0,313	1,748	0,065	1,376	0,499
$Mat$	0,001	0,139	0,001	0,353	0,001	0,222	0,001	0,513
$Age_S$	-0,010	0,027	-0,010	0,106	-0,020	0,000	-0,013	0,125
$Age_D$	0,021	0,092	0,022	0,217	0,007	0,514	-0,013	0,513
$DM$	0,045	0,000	0,050	0,000	0,060	0,000	0,067	0,000
$K$	-	-	1,888	0,000	-	-	1,288	0,000
Modelo 0 inflado								
$F_S$	6,016	0,672	6,761	0,085	-	-	-	-
$F_D$	2,189	0,672	3,368	0,534	-	-	-	-
$T_i$	-14,062	0,026	-15,526	0,032	-	-	-	-
$Mat$	-0,001	0,621	-0,001	0,647	-	-	-	-
$Age_S$	0,037	0,086	0,0354	0,121	-	-	-	-
$Age_D$	0,053	0,231	0,068	0,191	-	-	-	-
$DM$	-0,036	0,387	-0,056	0,102	-	-	-	-
AIC	629,733	-	607,24	-	780,95	-	-	-

En la tabla 3.2 se dividieron los datos en dos periodos: 59 familias que vivieron del 1400 a 1599 y 65 familias que vivieron del 1600 al 1800. Como en el análisis anterior, el modelo que mejor ajusto en los dos periodos fue el modelo ZINB, lo que denota una dispersión del número de hijos provocada tanto por una gran cantidad de ceros como por la propia dispersión de la variable. Los test de razón de verosimilitud entre el modelo ZIP y ZINB fueron altamente significativos en los dos periodos. Los resultados indican que en la etapa que va desde 1400 hasta 1599 no hubo ningún efecto de la consanguinidad sobre la fertilidad completada o la infertilidad de las parejas. Sin embargo, en el periodo que va desde 1600 al 1800 se detectó un efecto positivo significativo de la consanguinidad de los varones en el modelo de ceros inflados (coef = 33,73,  $P=0,034$ ). Esto quiere decir que la consanguinidad de los varones tuvo un efecto sobre la infertilidad de las parejas de los Habsburgo en el periodo 1600 – 1800. En cuanto a las variables nuisance, se puede apreciar como la duración del matrimonio fue de relevancia en los dos periodos; las

variables edad de los varones y las mujeres en la etapa 1600 – 1800. En el caso de la edad de los varones, tuvo un efecto negativo sobre el número de hijos de la pareja y en la caso de la edad de las mujeres tuvo un efecto positivo sobre la infertilidad de la pareja; es decir, en los dos casos estamos ante un ejemplo de senescencia en fertilidad y fecundidad.

**Tabla 3.2.** Análisis de regresión para todo el set de datos, clasificados según dos periodos (1400 – 1599 y 1600 – 1800), de la fertilidad completada – número de hijos – en función de los coeficientes de consanguinidad de los progenitores y coascendencia ( $F_S$ ,  $F_D$ ,  $T_i$ , respectivamente), la fecha del matrimonio ( $Mat$ ), la edad media de los varones ( $Age_S$ ), la edad media de las hembras ( $Age_D$ ) y la duración del matrimonio ( $DM$ ). Se utilizaron modelos GLM por Poisson (expresión 2.6), Binomial negativa (expresión 2.10), además de modelos mixturados ZIP (expresiones 2.16, 2.17) y ZINB y (expresiones 2.26, 2.27).  $K$  es el parámetro de dispersión de la Binomial negativa y el AIC es el criterio de información de Akaike.

	ZIP		ZINB		Poisson		Binom. Neg.	
	Esti.	P-val	Esti.	P-val	Esti.	P-val	Esti.	P-val
1400 - 1599								
Modelo por contaje								
$F_S$	-1,257	0,489	-1,629	0,520	1,489	0,400	0,943	0,825
$F_D$	4,564	0,028	4,421	0,137	2,291	0,207	4,113	0,335
$T_i$	-1,173	0,442	-1,180	0,577	1,625	0,279	0,601	0,896
$Mat$	0,003	0,091	0,003	0,144	0,003	0,136	0,004	0,404
$Age_S$	-0,001	0,879	-0,001	0,938	-0,011	0,054	-0,004	0,742
$Age_D$	0,017	0,334	0,008	0,760	0,012	0,436	-0,006	0,847
$DM$	0,047	0,000	0,049	0,000	0,076	0,000	0,082	0,000
$K$	-	-	1,914	0,000	-	-	1,135	0,000
Modelo 0 inflado								
$F_S$	-14,92	0,308	-17,19	0,331	-	-	-	-
$F_D$	-3,289	0,743	-2,631	0,797	-	-	-	-
$T_i$	-6,596	0,542	-6,683	0,573	-	-	-	-
$Mat$	-0,002	0,827	-0,001	0,903	-	-	-	-
$Age_S$	0,031	0,307	0,032	0,293	-	-	-	-
$Age_D$	0,026	0,704	0,014	0,862	-	-	-	-
$DM$	-0,122	0,000	-0,120	0,013	-	-	-	-
AIC	300,580	-	292,64	-	367,610	-	302,620	-
1600 - 1800								
Modelo por contaje								
$F_S$	1,055	0,407	1,302	0,458	-3,721	0,001	-4,322	0,039
$F_D$	1,455	0,329	1,826	0,337	-2,037	0,156	-1,259	0,446
$T_i$	-2,105	0,125	-2,238	0,213	1,198	0,353	2,233	0,349
$Mat$	0,002	0,144	0,002	0,344	-0,001	0,614	-0,001	0,540
$Age_S$	-0,022	0,001	-0,024	0,009	-0,029	0,000	-0,032	0,007
$Age_D$	0,045	0,011	0,047	0,044	-0,003	0,851	-0,018	0,481
$DM$	0,053	0,000	0,056	0,000	0,044	0,000	0,051	0,003
$K$	-	-	0,206	0,000	-	-	1,792	0,000
Modelo 0 inflado								
$F_S$	28,73	0,021	33,74	0,031	-	-	-	-
$F_D$	24,74	0,094	29,67	0,051	-	-	-	-
$T_i$	-27,14	0,049	-31,31	0,136	-	-	-	-
$Mat$	0,020	0,099	0,023	0,098	-	-	-	-
$Age_S$	0,059	0,179	0,064	0,008	-	-	-	-
$Age_D$	0,413	0,032	0,484	0,039	-	-	-	-
$DM$	0,100	0,248	0,118	0,13	-	-	-	-
AIC	329,11	-	320,87	-	404,11	-	344,16	-

### 3.2.2. Análisis de los datos correspondientes a las mujeres de 17 a 27 años de edad

En la tabla 3.3 se presentan los resultados obtenidos para 99 parejas con mujeres de 17 a 27 años de edad, lo que corresponde a mujeres que tuvieron hijos en su juventud. El modelo que mejor ajusta, según el AIC, es el modelo ZIP; tanto el parámetro  $k$  – del modelo ZINB – como el test de razón de verosimilitud entre los dos modelos de ceros inflados fueron no significativos ( $P=0.920$  y  $P=0.978$  respectivamente). En la figura 3.3 se puede apreciar como no se detecta ningún efecto de la consanguinidad sobre la fertilidad completada o la infertilidad de las parejas. Tan sólo se detectó un efecto significativo de la duración del matrimonio sobre el número total de hijos.

**Tabla 3.3.** Análisis de regresión, para los datos correspondientes al rango de edad de las hembras de 17 a 27 años, de la fertilidad completada – número de hijos – en función de los coeficientes de consanguinidad de los progenitores y coascendencia ( $F_s$ ,  $F_D$ ,  $T_i$ , respectivamente), la fecha del matrimonio ( $Mat$ ), la edad media de los varones ( $AgeS$ ) y la duración del matrimonio ( $DM$ ). Se utilizaron modelos GLM por Poisson (expresión 2.7), Binomial negativa (expresión 2.11), además de modelos mixturados ZIP (expresiones 2.19, 2.20) y ZINB (expresiones 2.28 y 2.29).  $K$  es el parámetro de dispersión de la Binomial negativa y el AIC es el criterio de información de Akaike.

	ZIP		ZINB		Poisson		Binom. Neg.	
	Esti.	P-val	Esti.	P-val	Esti.	P-val	Esti.	P-val
Modelo por contaje								
$F_s$	-0,903	0,485	-0,902	0,485	-1,785	0,099	-1,642	0,294
$F_D$	-0,027	0,983	-0,027	0,983	-0,991	0,439	-0,807	0,667
$T_i$	-0,245	0,854	-0,245	0,854	1,756	0,176	1,727	0,366
$Mat$	0,001	0,325	0,000	0,325	0,001	0,354	0,001	0,473
$AgeS$	-0,004	0,518	-0,004	0,518	-0,014	0,023	-0,124	0,170
$DM$	0,152	0,000	0,152	0,000	0,158	0,000	0,169	0,000
$K$	-	-	13,650	0,920	-	-	2,792	0,000
Modelo 0 inflado								
$F_s$	3,225	0,467	3,221	0,467	-	-	-	-
$F_D$	2,774	0,594	2,682	0,609	-	-	-	-
$T_i$	-10,17	0,131	-10,10	0,130	-	-	-	-
$Mat$	0,005	0,877	0,005	0,876	-	-	-	-
$AgeS$	0,004	0,183	0,036	0,182	-	-	-	-
$DM$	-0,005	0,654	-0,053	0,638	-	-	-	-
AIC	409,91	-	411,910	-	467,23	-	446,420	-

Tampoco se encuentra efecto significativo de los distintos coeficientes de consanguinidad sobre la variable respuesta cuando se dividen los datos en dos periodos 1400 – 1599 y 1600 – 1800 (tabla 3.4). Los resultados son idénticos a los obtenidos analizando todos los datos en un solo periodo 1400 – 1800; el modelo que mejor explica los datos es el modelo ZIP y tan sólo la duración del matrimonio ha tenido un efecto significativo sobre el número de hijos de las parejas Habsburgo con mujeres con edades comprendidas entre los 17 y los 27 años.

**Tabla 3.4.** Análisis de regresión, para los datos correspondientes al rango de edad de las hembras de 17 a 27 años – clasificados por dos épocas (1400 – 1599 y 1600 – 1800) – de la fertilidad completada – número de hijos – en función de los coeficientes de consanguinidad de los progenitores y coascendencia ( $F_S$ ,  $F_D$ ,  $T_i$ , respectivamente), la fecha del matrimonio ( $Mat$ ), la edad media de los varones ( $AgeS$ ) y la duración del matrimonio ( $DM$ ). Se utilizaron modelos GLM por Poisson (expresión 2.7), Binomial negativa (expresión 2.11), además de modelos mixturados ZIP (expresiones 2.19, 2.20) y ZINB (expresiones 2.28 y 2.29).  $K$  es el parámetro de dispersión de la Binomial negativa y el AIC es el criterio de información de Akaike.

	ZIP		ZINB		Poisson		Binom. Neg.	
	Esti.	P-val	Esti.	P-val	Esti.	P-val	Esti.	P-val
1400 - 1599								
Modelo por contaje								
$F_S$	-0,924	0,643	-0,886	0,660	1,517	0,455	1,565	0,589
$F_D$	1,958	0,379	1,951	0,381	-2,104	0,351	-1,556	0,598
$T_i$	0,207	0,921	0,173	0,934	3,034	0,167	3,045	0,296
$Mat$	0,003	0,232	0,003	0,231	0,003	0,278	0,003	0,457
$AgeS$	-0,013	0,183	-0,013	0,184	-0,025	0,008	-0,026	0,045
$DM$	0,182	0,000	0,182	0,000	0,213	0,000	0,221	0,000
$K$	-	-	11,30	0,814	-	-	4,58	0,037
Modelo 0 inflado								
$F_S$	-11,99	0,16	-12,000	0,160	-	-	-	-
$F_D$	6,94	0,502	6,955	0,489	-	-	-	-
$T_i$	-3,924	0,811	-4,053	0,805	-	-	-	-
$Mat$	-0,001	0,913	-0,001	0,914	-	-	-	-
$AgeS$	0,048	0,378	0,486	0,375	-	-	-	-
$DM$	-0,195	0,356	-0,196	0,313	-	-	-	-
AIC	190,63	-	192,630	-	212,19	-	209,83	-
1600 - 1800								
Modelo por contaje								
$F_S$	-1,391	0,450	-1,470	0,521	-3,352	0,023	-3,636	0,075
$F_D$	-1,104	0,550	-1,020	0,581	-1,187	0,506	-1,289	0,611
$T_i$	-0,417	0,819	-0,404	0,824	0,814	0,639	0,913	0,712
$Mat$	0,001	0,518	0,001	0,518	0,003	0,829	0,001	0,986
$AgeS$	0,008	0,432	0,007	0,478	-0,007	0,396	-0,004	0,712
$DM$	0,138	0,000	0,138	0,000	0,129	0,000	0,132	0,002
$K$	-	-	10,43	0,944	-	-	2,890	0,000
Modelo 0 inflado								
$F_S$	10,49	0,188	10,32	0,189	-	-	-	-
$F_D$	0,951	0,897	0,272	0,902	-	-	-	-
$T_i$	-6,509	0,481	-5,844	0,482	-	-	-	-
$Mat$	0,008	0,469	0,008	0,471	-	-	-	-
$AgeS$	0,059	0,288	0,058	0,289	-	-	-	-
$DM$	0,049	0,793	0,051	0,794	-	-	-	-
AIC	230,27	-	232,280	-	255,29	-	245,58	-

## 3.2.3. Análisis de los datos correspondientes a las mujeres de 30 a 40 años de edad

A diferencia de lo que sucedía con las parejas clasificadas según el rango de edad de 17 a 27 años, cuando se clasifican los datos según la edad de las mujeres de 30 a 40 años, se encuentra un efecto significativo y positivo de la consanguinidad sobre la infertilidad. En la tabla 3.5 se presentan los resultados de los análisis de 92 parejas con mujeres con edades comprendidas entre los 30 y los 40, es decir, mujeres que tuvieron descendencia en la última etapa de su periodo fértil. Se puede apreciar como la consanguinidad de los varones tiene un efecto positivo y significativo sobre la infertilidad (coef = 10.99  $P = 0.041$ ) modelizada en los ceros inflados. En el modelo por contaje se encuentra un efecto significativo positivo de la duración del matrimonio, así como el número de infantes nacidos en edades más tempranas de las madres. En el caso de la infertilidad – ceros inflados – la edad del varón y el número de infantes nacidos en edades más tempranas de las mujeres también tienen efectos significativos. En este caso, según el AIC, el modelo que mejor ajusta también es el ZIP, tanto el parámetro de sobredispersión del modelo ZINB como el test de verosimilitud – entre ZIP y ZINB – fueron no significativos ( $P = 0.827$  y  $P = 0.712$  respectivamente).

**Tabla 3.5.** Análisis de regresión, para los datos correspondientes al rango de edad de las hembras de 30 a 40 años, de la fertilidad completada – número de hijos – en función de los coeficientes de consanguinidad de los progenitores y coascendencia ( $F_S$ ,  $F_D$ ,  $T_i$ , respectivamente), la fecha del matrimonio (Mat), la edad media de los varones (AgeS), niños nacidos con anterioridad (N), mortalidad infantil en la etapa anterior (M) y la duración del matrimonio (DM). Se utilizaron modelos GLM por Poisson (expresión 2.8), Binomial negativa (expresión 2.12), además de modelos mixturados ZIP (expresiones 2.21, 2.22) y ZINB (expresiones 2.30 y 2.31). K es el parámetro de dispersión de la Binomial negativa y el AIC es el criterio de información de Akaike.

	ZIP		ZINB		Poisson		Binom. Neg.	
	Esti.	P-val	Esti.	P-val	Esti.	P-val	Esti.	P-val
Modelo por contaje								
$F_S$	-1,890	0,302	-1,894	0,409	-4,400	0,011	-5,386	0,049
$F_D$	-0,842	0,704	-0,847	0,708	-1,437	0,510	-1,218	0,734
$T_i$	2,193	0,193	2,192	0,194	4,420	0,007	4,240	0,122
Mat	0,004	0,600	0,004	0,600	0,003	0,659	-0,001	0,943
AgeS	-0,001	0,881	-0,001	0,881	-0,030	0,000	-0,022	0,047
DM	0,182	0,000	0,182	0,000	0,179	0,000	0,157	0,018
N	0,074	0,011	0,074	0,011	0,1732	0,000	0,169	0,000
M	-0,093	0,252	-0,093	0,252	-0,089	0,269	0,015	0,918
K	-	-	11,102	0,914	-	-	1,407	0,000
Modelo 0 inflado								
$F_S$	10,99	0,041	10,95	0,040	-	-	-	-
$F_D$	-0,827	0,920	-0,919	0,921	-	-	-	-
$T_i$	-9,147	0,176	-9,102	0,175	-	-	-	-
Mat	-0,004	0,914	-0,004	0,914	-	-	-	-
AgeS	0,091	0,002	0,091	0,002	-	-	-	-
DM	0,179	0,288	0,178	0,288	-	-	-	-
N	-0,528	0,004	-0,527	0,004	-	-	-	-
M	0,020	0,964	0,020	0,963	-	-	-	-
AIC	328,96	-	330,96	-	389,13	-	365,96	-

Cuando se dividen los datos en dos periodos 1400 – 1599 y 1600 – 1800 (42 y 40 parejas respectivamente) se puede apreciar, otra vez, cómo el efecto de la consanguinidad de los varones sobre la infertilidad de las parejas es positiva y significativa en la segunda etapa: 1600 – 1800 (coef = 381.7,  $P = 0.0480$ ; tabla 3.6). El modelo ZIP fue el que mejor ajustó los datos en los dos periodos. En cuanto a las variables nuisance, cabe destacar que en la etapa 1600 – 1800, en el modelo por contaje – fertilidad

completada – la edad del varón, el número de infantes nacidos en edades más tempranas de las mujeres (N) y el número de muertes infantiles en edades más tempranas de las mujeres (M) tuvieron efectos significativos.

**Tabla 3.6.** Análisis de regresión, para los datos correspondientes al rango de edad de las hembras de 30 a 40 años – clasificados por dos épocas (1400 – 1599 y 1600 – 1800) – de la fertilidad completada, en función de los coeficientes de consanguinidad de los progenitores y coascendencia ( $F_S$ ,  $F_D$ ,  $T_i$ , respectivamente), la fecha del matrimonio (Mat), la edad media de los varones (AgeS), niños nacidos con anterioridad (N), mortalidad infantil en la etapa anterior (M) y la duración del matrimonio (DM). Se utilizaron modelos GLM por Poisson (expresión 2.8), Binomial negativa (expresión 2.12), además de modelos mixturados ZIP (expresiones 2.21, 2.22) y ZINB (expresiones 2.30 y 2.31). K es el parámetro de dispersión de la Binomial negativa y el AIC es el criterio de información de Akaike

	ZIP		ZINB		Poisson		Binom. Neg.	
	Esti.	P-val	Esti.	P-val	Esti.	P-val	Esti.	P-val
1400 - 1600								
Modelo por contaje								
$F_S$	-4,637	0,349	-4,664	0,349	-5,324	0,220	-4,655	0,591
$F_D$	2,273	0,604	2,260	0,604	0,014	0,997	-2,900	0,710
$T_i$	3,281	0,313	3,311	0,313	4,635	0,093	2,626	0,685
Mat	-0,004	0,284	-0,004	0,284	0,000	0,941	0,003	0,683
AgeS	0,012	0,122	0,012	0,122	-0,009	0,281	-0,007	0,672
DM	0,266	0,002	0,267	0,002	0,187	0,028	0,128	0,301
N	-0,002	0,961	-0,002	0,960	0,137	0,006	0,181	0,081
M	0,038	0,783	0,038	0,783	0,001	0,992	0,279	0,368
K	-	-	10,114	0,813	-	-	0,834	0,000
Modelo 0 inflado								
$F_S$	3,973	0,839	3,712	0,833	-	-	-	-
$F_D$	5,245	0,711	5,888	0,699	-	-	-	-
$T_i$	-7,328	0,633	-7,133	0,629	-	-	-	-
Mat	-0,010	0,499	-0,011	0,499	-	-	-	-
AgeS	0,056	0,128	0,055	0,124	-	-	-	-
DM	0,256	0,314	0,264	0,313	-	-	-	-
N	-0,448	0,048	-0,450	0,048	-	-	-	-
M	-0,355	0,656	-0,371	0,653	-	-	-	-
AIC	159,84	-	161,84	-	197,48	-	172,23	-
1600 - 1800								
Modelo por contaje								
$F_S$	-0,372	0,861	-0,324	0,829	-5,260	0,020	-5,756	0,027
$F_D$	-4,291	0,197	-4,505	0,141	-3,669	0,297	-3,983	0,314
$T_i$	1,354	0,574	1,467	0,533	5,015	0,029	5,891	0,027
Mat	0,000	0,762	0,000	0,765	-0,003	0,150	-0,003	0,158
AgeS	-0,039	0,000	-0,038	0,001	-0,060	0,000	-0,061	0,000
DM	0,122	0,072	0,120	0,014	0,227	0,000	0,212	0,004
N	0,107	0,005	0,108	0,004	0,199	0,000	0,207	0,000
M	-0,335	0,005	-0,337	0,004	-0,199	0,082	-0,212	0,110
K	-	-	29,722	0,000	-	-	9,000	0,390
Modelo 0 inflado								
$F_S$	381,7	0,0480	369,8	0,052	-	-	-	-
$F_D$	-360,8	0,272	-341,3	0,332	-	-	-	-
$T_i$	-99,48	0,429	-101,1	0,459	-	-	-	-
Mat	0,313	0,216	0,303	0,311	-	-	-	-
AgeS	1,780	0,339	1,730	0,404	-	-	-	-
DM	-2,926	0,226	-2,711	0,228	-	-	-	-
N	-5,999	0,250	-5,834	0,362	-	-	-	-
M	-25,53	0,206	-24,69	0,349	-	-	-	-
AIC	161,74	-	167,74	-	194,83	-	196,1	-

### 3.3. Efectos de la consanguinidad sobre el periodo reproductivo corregido

Además de estudiar el efecto de los matrimonios entre parientes sobre la fertilidad a través del número de hijos de las parejas Habsburgo – fertilidad completada – también se utilizó el periodo reproductivo corregido de las parejas como otra variable respuesta de la fertilidad sobre la que evaluar el efecto de los coeficientes de consanguinidad. Se definió, de esta manera, el periodo reproductivo corregido (PRC) como el tiempo, en años, transcurrido entre el nacimiento del primer hijo y el último, a esta diferencia se le suma uno.

En la tabla 3.7 se presentan los resultados obtenidos para el análisis de los modelos, presentados en los materiales y métodos, de 124 parejas Habsburgo en el periodo 1400 – 1800. En la tabla, se puede apreciar como el modelo ZINB es el que mejor ajusta, ya que tiene el menor AIC, así mismo, el parámetro de sobredispersión  $k$  es altamente significativo ( $P < 0.000$ ) al igual que el test de razón de verosimilitud entre el modelo ZIP y ZINB ( $P = 0.001$ ). No se detecta ningún efecto significativo de los distintos coeficientes de consanguinidad sobre el periodo reproductivo corregido en el moldeo por contagio, es decir, la fertilidad. Sin embargo, hay un efecto negativo significativo de la consanguinidad de la descendencia (o coascendencia de la pareja) sobre la infertilidad. Este efecto se había encontrado también en el análisis de la fertilidad completada con los 124 datos. Es interesante comentar que de las variables nuisance tanto la edad de la mujer como la duración del matrimonio, en el modelo por contagio, son significativas, lo que resulta lógico. La duración del matrimonio también tuvo un efecto negativo significativo sobre la infertilidad de las parejas.

En la tabla 3.8 se presentan los resultados para los datos divididos en dos etapas distintas: 1400 – 1599 y 1600 – 1800 (59 y 65 parejas respectivamente). Como en el caso anterior, el modelo ZINB es el que mejor ajusta en los dos periodos; los parámetros de sobredispersión, así como los test de razón de verosimilitud fueron significativos en los dos periodos. En cuanto al efecto de los coeficientes de consanguinidad, se puede apreciar como en la etapa 1600 – 1800 se encuentra un efecto positivo, significativo, de la consanguinidad de los varones sobre el modelo de ceros inflados del periodo reproductivo corregido (coef = 22.39,  $P = 0.018$ ). Este efecto también se encontró en este mismo periodo utilizando como medida de fertilidad el número de hijos de cada pareja. Esto tiene todo el sentido, ya que las familias con cero hijos son las mismas se haya utilizado como variable respuesta la fertilidad completada o el periodo reproductivo corregido. En el periodo de 1600 – 1800 se puede observar que tanto la edad de la mujer, como del varón, así como la duración del matrimonio tienen un efecto significativo sobre el periodo reproductivo corregido. En el caso de la edad del varón este efecto es de signo negativo, lo que denota un efecto de la senescencia masculina en la fecundidad de la pareja.

**Tabla 3.7.** Análisis de regresión para todo el set de datos (n=124) del periodo reproductivo corregido (PRC) en función de los coeficientes de consanguinidad de los progenitores y coascendencia ( $F_s$ ,  $F_D$ ,  $T_i$ , respectivamente), la fecha del matrimonio (Mat), la edad media de los varones (AgeS), la edad media de las hembras (AgeD) y la duración del matrimonio (DM). Se utilizaron modelos GLM por Poisson (expresión 2.9), Binomial negativa (expresión 2.13), además de modelos mixturados ZIP (expresiones 2.23, 2.24) y ZINB (expresiones 2.32 y 2.33). K es el parámetro de dispersión de la Binomial negativa y el AIC es el criterio de información de Akaike.

	ZIP		ZINB		Poisson		Binom. Neg.	
	Esti.	P-val	Esti.	P-val	Esti.	P-val	Esti.	P-val
Modelo por contaje								
$F_s$	-0,038	0,962	0,008	0,952	-1,974	0,004	-2,037	0,311
$F_D$	1,434	0,114	1,99	0,206	0,273	0,751	1,528	0,527
$T_i$	-0,604	0,445	-1,018	0,463	2,457	0,001	2,265	0,315
Mat	0,008	0,044	0,000	0,339	0,000	0,216	0,004	0,731
AgeS	-0,011	0,001	-0,010	0,071	-0,022	0,000	-0,014	0,138
AgeD	0,037	0,000	0,041	0,018	0,019	0,021	-0,003	0,880
DM	0,051	0,000	0,059	0,000	0,066	0,000	0,081	0,000
K	-		1,597	0,000	-	-	0,903	0,000
Modelo 0 inflado								
$F_s$	5,835	0,158	6,375	0,139	-	-	-	-
$F_D$	1,715	0,736	2,371	0,622	-	-	-	-
$T_i$	-14,23	0,019	-11,530	0,023	-	-	-	-
Mat	0,000	0,992	0,000	0,970	-	-	-	-
AgeS	0,041	0,052	0,045	0,071	-	-	-	-
AgeD	0,052	0,219	0,067	0,112	-	-	-	-
DM	0,075	0,019	-0,061	0,047	-	-	-	-
AIC	749,388		687,21		1057,8		689,640	

Tabla 3.8. Análisis de regresión para todo el set de datos (n=124) – clasificados por dos épocas (1400 – 1599 y 1600 – 1800) – del periodo reproductivo corregido (PRC) en función de los coeficientes de consanguinidad de los progenitores y coascendencia ( $F_s$ ,  $F_D$ ,  $T_i$ , respectivamente), la fecha del matrimonio (Mat), la edad media de los varones (AgeS), la edad media de las hembras (AgeD) y la duración del matrimonio (DM). Se utilizaron modelos GLM por Poisson (expresión 2.9), Binomial negativa (expresión 2.13), además de modelos mixturados ZIP (expresiones 2.23, 2.24) y ZINB (expresiones 2.32 y 2.33). K es el parámetro de dispersión de la Binomial negativa y el AIC es el criterio de información de Akaike.

	ZIP		ZINB		Poisson		Binom. Neg.	
	Esti.	P-val	Esti.	P-val	Esti.	P-val	Esti.	P-val
1400 - 1599								
Modelo por contaje								
$F_s$	-0,632	0,670	-0,755	0,792	1,361	0,360	0,522	0,914
$F_D$	6,110	0,000	6,624	0,056	3,481	0,017	5,975	0,214
$T_i$	-0,198	0,132	-2,053	0,393	1,167	0,359	0,609	0,883
Mat	0,000	0,618	0,001	0,685	0,000	0,987	0,001	0,789
AgeS	-0,001	0,951	-0,001	0,984	-0,010	0,024	-0,004	0,776
AgeD	0,041	0,002	0,043	0,130	0,029	0,018	0,012	0,720
DM	0,055	0,000	0,063	0,000	0,080	0,000	0,093	0,000
K			1,280	0,000	-	-	0,787	0,023
Modelo 0 inflado								
$F_s$	-9,159	0,497	-9,331	0,510	-	-	-	-
$F_D$	0,408	0,966	1,335	0,893	-	-	-	-
$T_i$	-10,78	0,377	-10,97	0,382	-	-	-	-
Mat	-0,003	0,683	-0,003	0,735	-	-	-	-
AgeS	0,028	0,324	0,029	0,317	-	-	-	-
AgeD	0,051	0,440	0,057	0,422	-	-	-	-
DM	-0,096	0,028	-0,090	0,0491	-	-	-	-
AIC	373,21		335,62		509,34		344,08	-
1600 - 1800								
Modelo por contaje								
$F_s$	0,373	0,740	0,359	0,829	-3,773	0,000	-5,139	0,021
$F_D$	0,959	0,484	1,374	0,427	-1,803	0,111	-1,937	0,472
$T_i$	-0,084	0,939	-0,263	0,865	3,130	0,001	4,588	0,075
Mat	0,001	0,199	0,000	0,521	-0,001	0,252	-0,002	0,371
AgeS	-0,023	0,000	-0,026	0,000	-0,032	0,000	-0,037	0,003
AgeD	0,051	0,001	0,056	0,004	0,009	0,406	-0,008	0,744
DM	0,056	0,000	0,062	0,000	0,054	0,000	0,069	0,000
K	-	-	2,157	0,000	-	-	-	-
Modelo 0 inflado								
$F_s$	19,13	0,065	22,39	0,018	-	-	-	-
$F_D$	12,12	0,456	16,13	0,187	-	-	-	-
$T_i$	-20,43	0,271	-24,17	0,062	-	-	-	-
Mat	0,016	0,102	0,018	0,071	-	-	-	-
AgeS	0,068	0,098	0,065	0,110	-	-	-	-
AgeD	0,235	0,227	0,294	0,043	-	-	-	-
DM	-0,001	0,982	0,015	0,806	-	-	-	-
AIC	375,79		362,43		534,54		-	-

## ● Conclusiones y discusión



En este trabajo de fin de máster se han utilizado los modelos GLM y ceros inflados para estudiar el efecto de la consanguinidad de las dinastías Habsburgo en su propia fertilidad, medida a través de dos variables: número de hijos de las parejas y periodo reproductivo de las mujeres. Es la primera vez que se utilizan los modelos de ceros inflados en el estudio de la depresión consanguínea en fertilidad humana, ya que hasta ahora en la mayoría de los estudios que se ocupaban de esta problemática, utilizando el número de hijos como variable dependiente de fertilidad, ignoraban aquellas parejas sin hijos (Robert et al., 2009; Postma et al., 2010). La aplicación de los ceros inflados permite modelizar a la vez la relación entre la consanguinidad y el número de hijos de las parejas – depresión consanguínea en fertilidad – además de la relación entre la consanguinidad y la infertilidad de la pareja. De modo general, se puede concluir que en la dinastía Habsburgo, en determinadas ocasiones, se ha podido detectar un efecto de la consanguinidad de los varones en la infertilidad de las parejas Habsburgo. Así mismo, se ha detectado efecto de la senescencia de varones y mujeres en la fertilidad de las parejas.

Para analizar la depresión consanguínea en fertilidad medida esta a través del número de hijos de cada pareja se dividieron los datos en distintos conjuntos: un primer conjunto con todos los datos – 124 familias – y posteriormente en dos conjuntos según la edad de las madres: de 17 a 27 años y de 30 a 40 años. De esta forma se pretendió analizar el efecto de la senescencia materna en el número de hijos de las parejas. Esta división de los datos se ha visto lógica, ya que el modelo que mejor ha ajustado el conjunto de todos los datos fue el ZINB, mientras que cuando se analizaron los conjuntos según la edad de las madres el modelo que mejor ajusto en los dos casos fue el ZIP. Es decir, cuando se analizan todos los datos en su conjunto, además de haber más ceros de los esperados, se produce sobre-dispersión debido a subestructura poblacional que es eliminada al dividir los datos según la edad de las mujeres. Los dos conjuntos creados según la edad de las madres son más homogéneos y por lo tanto el modelo de regresión que mejor ajusta es aquel con ceros inflados pero sin sobre-dispersión. En cuanto a los efectos de la consanguinidad, no se ha podido detectar bajo ningún modelo depresión consanguínea en fertilidad completada, ni por el coeficiente de consanguinidad de varones, mujeres o la propia descendencia. Para explicar este hecho se podría apuntar, por ejemplo, a posibles eventos de compensación reproductora, sin embargo, hace falta más análisis para poder concluir algo con respecto a este tema. En cambio, en dos conjuntos de datos, se ha detectado un efecto de la consanguinidad de los varones sobre la infertilidad de las parejas. Este efecto sobre la infertilidad se ha detectado en el conjunto de todos los datos, pero tan sólo en el segundo periodo de 1600 a 1800. Este resultado sugiere un proceso subyacente de anti-purgado, es decir, durante generaciones ha habido acumulación en los varones de genes deletéreos con efectos muy suaves – y por lo tanto indetectables por la selección natural – que llegado un punto, con el aumento de la consanguinidad, provocan un efecto sobre la infertilidad de las parejas. Cuando se divide el conjunto de datos según la edad de las madres se puede ver el efecto de la senescencia combinado con el efecto de la consanguinidad. Mientras que en el grupo de edad de las madres 17 a 27 años no se detecta ningún efecto de la consanguinidad, en el grupo de edad de las madres de 30 a 40 se puede observar un efecto significativo y positivo del coeficiente de consanguinidad de los varones sobre la infertilidad de las parejas; tanto en todo el periodo de estudio (1400 – 1800) como en la última etapa (1600 – 1800), no así en la primera etapa (1400 – 1599) lo que indica, otra vez, un posible evento de anti-purgado. Estos resultados sugieren que en las mujeres mayores, los efectos de la consanguinidad de los varones sobre la infertilidad de las parejas son mayores. Estos efectos de la senescencia en la fertilidad completada se han detectado en prácticamente todos los análisis realizados para este trabajo, ya que en muchos de ellos, tanto la edad de los padres como de las madres, tienen

efectos significativos sobre la fertilidad completada. Es decir, cuanto mayor sea el padre o la madre (dentro del periodo reproductivo natural de la madre) menor será el número de hijos que den a luz. Para terminar comentar que no hubo efecto del año del matrimonio en ningún modelo, por lo que, como se ha sugerido en estudios anteriores (Ceballos y Álvarez, 2013) se puede suponer que no hay mucha heterogeneidad ambiental a lo largo de los 400 años que se extiende este estudio, o que si la hay, ésta prácticamente no tiene efecto sobre la fertilidad completada.

Los resultados arrojados por el análisis sobre el periodo reproductivo corregido son muy parecidos a los obtenidos con la variable número de hijos por pareja. No se ha detectado un efecto significativo de la consanguinidad sobre la fertilidad de las 124 parejas Habsburgo. El modelo que mejor ajusta en este caso fue el ZINB. Los resultados de los modelos de ceros inflados son equivalentes a los obtenidos en los análisis de las tablas 3.1 y 3.2, esto es así porque las familias de cero hijos son las mismas para los análisis con las dos variables de fertilidad. De esta manera, se encontró un efecto significativo del coeficiente de consanguinidad de los varones en la infertilidad de las parejas en el periodo 1600 – 1800.

A pesar de que la depresión consanguínea en supervivencia pre-reproductiva en los Habsburgo fue de gran intensidad y fácil de detectar (Ceballos y Álvarez, 2013), no ha sucedido lo mismo con el otro parámetro de la eficacia biológica, la fertilidad. Este fracaso a la hora de detectar depresión consanguínea en fertilidad se debe a que esta variable depende de factores sociales y otros tipos de condicionantes que poco tienen que ver con la biología de los individuos y parejas. La ausencia de depresión consanguínea en los Habsburgo también podría explicarse por un proceso de compensación reproductora. Los Habsburgo, una de las familias más poderosa de la Europa de la época, estaban presionados socialmente para engendrar el mayor número de descendientes. Las familias más consanguíneas sufrían mayor mortalidad infantil (0 – 1 años) y pre-juvenil (1 – 10 años), por lo que debían tener un mayor número de hijos y extender sus periodos reproductivos para alcanzar el tamaño familiar socialmente aceptado. Por otra parte, la utilización de modelos mixturados con ceros inflados ha permitido encontrar un efecto del coeficiente de consanguinidad sobre la fertilidad de las parejas y además se ha encontrado que este efecto depende de la senescencia de las mujeres. Bajo nuestro conocimiento esta es la primera vez que se encuentra este efecto en el género humano.

Es necesario seguir realizando estudios para intentar limpiar los efectos de la consanguinidad sobre la fertilidad y de esta manera poder detectarlos. De esta manera, sería interesante realizar un estudio sobre el intervalo entre partos de las parejas Habsburgo utilizando modelos de datos de supervivencia, así como de un análisis de correlación entre el número de hijos de sucesivas generaciones, intentado conocer como fue la heredabilidad de esta variables de fertilidad en los Habsburgo.

# ● Bibliografía



- Alvarez, V.A. 2002. *Historia universal de la Edad Media*. Barcelona: Ariel Historia. 911 pp
- Alvarez, G., Ceballos, F.C y Quinteiro, C. 2009. The Role of Inbreeding in the Extinction of a European Royal Dynasty. *PLoS ONE*. 4(4): e5174. doi:10.1371/journal.pone.0005174
- Alvarez, G., Quinteiro C. y Ceballos, F.C. 2011. Inbreeding and Genetic Disorder.; en Ikehara, K. (Ed.). *Advances in the Study of Genetic Disorders*. InTech. DOI: 10.5772/745. 472 pp.
- Asa, C., Miller, P., Agnew, M., Rebolledo, J.A.R., Lindsey, S.L., Callahan, M. et al. 2007. Relationship of inbreeding with sperm quality and reproductive success in Mexican gray wolves. *Anim. Conserv.* 10: 326-331.
- Cavalli-Sforza, L.L. y Bodmer, W.F. 1971. *The Genetics of Human Population*. Freeman. San Francisco. 964 pp
- Ceballos, F.C. y Álvarez, G. 2013. Royal dynasties as human inbreeding laboratories: the Habsburgs. *Heredity* 111: 114-121.
- Charlesworth, B. 1994. *Evolution in Age-Structured populations*. Cambridge: Cambridge University Press. 324 pp
- Charlesworth, B. y Charlesworth, D. 1999. The genetic Basis of Inbreeding Depression. *Genet. Res.* 74: 329-40
- Charlesworth, D. y Willis, J.H. 2009. The Genetics of Inbreeding Depression. *Nat. Rev. Genet.* 10: 783-96
- Cotterman, C. W. 1940. *A Calculus for Statistico-genetics*. Unpublished thesis, Ohio State University. Columbus, Ohio.
- Crow, J.F., y Kimura, M. 1970. *An Introduction to Population Genetics Theory*. Harper and Row, New York
- Edwards, A.W. 1994. The fundamental theorem of natural selection. *Biol. Rev.* 69: 443-474
- Ballou, J.D. 1997. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J. Heredity*. 88: 169-78.
- Balloux, F., Amos, W. y Coulson, T. 2004. Does Heterozygosity Estimate Inbreeding in Real Populations?. *Mol. Ecol.* 13: 3021-31
- Bittles, A.H. y Neel, J.V. 1994. The cost of human inbreeding and their implications for the variations at the DNA level. *Nature Genetics*. 8: 117-21
- Bittles, A.H., Grant, J.C., Sullivan, S.G. y Hussain, R. 2002. Does inbreeding lead to decreases human fertility? *Ann. Hum. Biol.* 29: 111-130.
- Bittles, A.H. y Black, M.L. 2010a. The impact of consanguinity in neonatal and infant death. *Ear H. Develop.* Doi:10.1016/j.earhumdev.2010.08.003
- Bittles, A.H. y Black, M.L. 2010b. Consanguinity, Human evolution and complex diseases. *Proc. Natl. Acad. Sci. USA*. 107: 1779-86
- Bittles, A.H. 2012. *Consanguinity in Context*. Cambridge University Press. Cambridge. 318 pp
- Boyce, A.J. 1983. Computation of inbreeding and kinship coefficients on extended pedigrees. *J. Heredity*. 74: 400-4
- Brass, W. 1958. The Distribution of Births in Human Populations. *Popul. Stud.* 12: 57-72

## Bibliografía

- Broekmans, F.J., Faddy, M.J., Scheffer, G., y Velde, E.R. 2004. Antral follicle counts are related to age at natural fertility loss and age at menopause. *Menopause*. 11: 607-614
- Encyclopaedia Britannica, Inc (2007) *The New Encyclopaedia Britannica*, 15<sup>th</sup> edition. The House of Habsburg. Macropaedia. Chicago: Encyclopaedia Britannica, Inc
- Falconer, D.S., y Mackay T.F.C. 1996. *Introduction to Quantitative Genetics*. Pearson Prentice Hall. London. 480 pp
- Fisher, R.A. 1930. *The genetical theory of natural selection*. Oxford, England: Clarendon Press. 304 pp.
- Gargantilla, P. 2005. *Enfermedades de los reyes de España. Los Austrias: de la locura de Juana a la impotencia de Carlos II el hechizado*. Madrid, La Esfera de los Libros. 499 pp
- Geneall Project. 2008. Información genealógica de los países Europeos. Disponible en: <http://www.geneall.net>. Último acceso: 10 Junio 2014.
- Haldane, J.B.S. y Moshinsky, P. 1939. Inbreeding in mendelian populations with special reference to human cousin marriage. *Annals of Eugenics*. 9: 321-40
- Hawass, Z., Gad, Y.Z., Ismail, S., Khairat, R., Fathalla, D., et al. 2010. Ancestry and pathology in king Tutankhamun's family. *JAMA*. 303, 638-47.
- Helgason, A., Pálsson, S., Gudbjartsson, D.F., Kristjánsson, P. y Stefánsson, K. 2008. An association between the kinship and fertility of human couples. *Science* 319: 813-816.
- Hedrick, P.W. 2011. *Genetics of populations*. Jones and Bartlett, Sudbury.
- Hilbe, J.M. 2011. *Negative Binomial regression*. Cambridge University Press; 2 Edition. Cambridge. 576 pp.
- Houle, D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics*. 130:195-204
- Hughes, K.A., Alipaz J.A., Drnevich, J.M. y Reynolds, R.M. 2002. A test of evolutionary theories of aging. *Proc. Natl. Acad. Sci. USA*. 99: 14286-14291
- Imaizumi, Y. y Nei, M. 1970. Variability and heritability of human fertility. *Ann. Hum. Genet. Lond.* 33: 251-59.
- Jacquard, A. 1975. Inbreeding: One word, several meanings. *Theoretical Population Biology*. 7: 338-63
- Junceda, E. 1991. *Ginecología y Vida Íntima de las Reinas de España. Tomo I. Temas de Hoy*. Madrid. 278 pp
- Kimura, M. y Weiss, G.H. 1965. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*. 32: 237-250.
- Kümmel, W.F. 1990. *De Morbis Aulicis: On Diseases Found at Court.*; en: Nutton, V. (Ed.). *Medicine At The Courts of Europe, 1500 – 1837*. Routledge. London. 301 pp
- Lynch, J. 1991. *Spain, 1516-1700: From a nation state to world empire*. Basil Blackwell. Oxford. 514pp
- MacCluer, J.W., Boyce, A.J., Dyke, B., Weitkamp, L.R., Pfennig, D.W., et al. 1983. Inbreeding and pedigree structure in Standard bred horses. *J. Heredity*. 74: 394-9

- McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M. et al. 2008. Runs of Homozygosity in European Populations. *Am. J. Hum. Genet.* 83: 359-72
- Malécot, G. 1948. *Les mathématiques de L'hérédité*. Masson et Cie. Paris. 63 pp
- Malécot, G. 1966. *Probabilités et Hérédité*. Presses Universitaires de France. Paris. 357 pp
- Margulis, S.W. 1996. The effects of inbreeding on parental behavior and reproductive success in two species of the old-field mouse (*Peromyscus polionotus*). PhD thesis, Committee on Evolutionary Biology, University of Chicago, Chicago
- Morton, N.E., Crow, F.J. y Muller, H.J. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Natl. Acad. Sci. USA.* 42, 855-63
- Morton, N. E. 1977. Isolation by distance in human populations. *Ann. Hum. Genet.* 40: 361–365.
- Nelder, J y Wedderburn, J.R. 1972. Generalized Linear Models. *Jr. Stat. Soc.* 134: 370-384
- Nichols, S.M. Bavister, B.D., Brenner, C.A., et al. 2005. Ovarian senescence in the rhesus monkey (*Macaca mulatta*). *Hum. Reprod.* 20:79-83
- Labouriau, R. y Amorin, A. 2008. Human fertility increase with marital radius. *Genetics.* 178: 601-603
- Labouriau, R. y Amorin, A. 2008. Comment on “An association between the kinship and fertility of human couples”. *Science.* 322: 1634-1635
- Lloyd- Smith, J.O. 2007. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLoS ONE* 2 (2): e180. doi:10.1371/journal.pone.0000180
- Ober, C., Hyslop, T. y Hauck, W.W. 1999. Inbreeding effects on fertility in humans: evidence for reproductive compensation. *Am. J. Hum. Genet.* 64: 225-231.
- Peller, S. 1965. Births and deaths among europe's ruling families since 1500.; en: Glass, D.V. & Eversley, D.E.C. (Eds.). *Population in History*. London Edward Arnold LTD. London. 688 pp
- Pluzhnikov, A., Nolan, A.K., Tan, Z. y Ober, C. 2007. Correlation of intergenerational family sizes suggest a genetic component of reproductive fitness. *Am. J. Hum. Genet.* 81:165-169
- Postma, E., Martini, L. y Martini, P. 2010. Inbred women in a small Swiss village have fewer children. *J. Evol. Biol.* 23: 1468-1474.
- PRB. 2011. *World Population Data Sheet*. Washington, DC: Population Reference Bureau
- Promislow, D.E. y Tatar, M. 1998. Mutation and senescence: where genetics and demography meet. *Genetica.* 102: 299-314
- R Development Core Team. 2011. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org>.
- Ricklefs, R.E., Scheuerlein, A. y Cohen, A. 2003. Age-related patterns of fertility in captive populations of birds and mammals. *Exp. Geront.* 38: 741-745.
- Robert, A., Toupance, B., Tremblay, M. y Heyer, E. 2009. Impact of inbreeding on fertility in a pre-industrial population. *Eur. J. Hum. Genet.* 17: 673-681.
- Rodgers, J.L., Hughes, K., Kohler, H.T. et al. 2001. Genetic influence helps to explain variation in human fertility: Evidence from recent behavioural and molecular genetics studies. *Cur. Direct. Psych. Sci.* 10: 184-187

## Bibliografía

- Rodgers, J.L., Kohler, H.P. y Christensen, K. 2001. Behaviour genetic modeling of human fertility: Findings from a contemporary Danish twin study. *Demography*. 38:29-42
- Roldan, E.R.S., Cassinello, J., Abaigar, T. y Gomendio, M. 1998. Inbreeding, fluctuating asymmetry and ejaculate quality in an endangered ungulate. *Proc. R. Soc. B* 265: 243-248.
- Schierup, M. H. y Christiansen, F. B. 1996. Inbreeding depression and outbreeding depression in plants. *Heredity*, 77: 461–468.
- Seather, S.J. 2004. Density Estimation. *Stat. Sci.* 4: 588-597
- Simonoff, J.S. 1996. *Smoothing Methods in Statistics*. Springer
- Snoke, MS. y Promislow, D.E. 2003. Quantitative genetic test of recent senescence theory: age-specific mortality and male fertility in *Drosophila melanogaster*. *Heredity*. 69: 209-225
- Sutter, J. y Tabah, L. 1952. Effects de la consanguinité et de L'endogamie una enquête en morbihan et Loir-et-Cher. *Population*. 7: 511-26.
- Templeton, A.R. y Read, B. 1994. Inbreeding: One word, several meanings much confusion. *Conserv. Genet.* 68: 91-105
- Tetushkin, E. Y. 2011. Genetic Aspects of Genealogy. *Russian Journal of Genetics*. 47: 1288-306
- Vogel, F. y Motulsky. A.G. 1997. *Human Genetics. Problems and Approaches*. Springer-Verlag Berlin and Heidelberg GmbH &Co. K. New York
- Wand, M.P. y Jones, M.C. 1995. *Kernel Smoothing*. Chapman and Hall
- White, G.C. y Bennetts, R.E. 1996. Analysis of frequency count data using the negative binomial distribution. *Ecology* 77: 2549-2557.
- Williams, G.C. 1957. Pleiotropy, natural selection, and evolution of senescence. *Evolution*. 11: 398-411
- Wright, S. 1922. Coefficients of inbreeding and relationship. *American Naturalist*. 56: 330-38
- Wright, S. 1969. *Evolution and Genetics of Populations. II. The Theory of Gene Frequencies*. University of Chicago Press, Chicago. 520 pp
- Wood, J.W. 1994. *Dynamics of Human Reproduction. Biology, Biometry and Demography*. Aldine De Gruyter. New York. 651 pp
- Zeileis, A., Kleibe, C. & Jackman, S. 2008. Regression models for count data in R. *J. Stat. Softw.* 27: 1-25.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. 2009. *Mixed effects models and extensions in ecology with R*. Springer Science, New York.



