

Desafío teórico avanzado ISFG-GHEP 2025: parentesco

Diseñado por: Magnus Dehli Vigeland

Instrucciones generales

Esta es una prueba de opción múltiple que consta de 20 preguntas. Para cada pregunta, sólo una alternativa es correcta. Puede usar el software que desee, pero tenga en cuenta que algunos programas poseen convenciones integradas (por ejemplo, redondeo) que pueden afectar el resultado. Si su respuesta no coincide exactamente con ninguna de las opciones, elija la más cercana.

Suposiciones: A lo largo de la prueba, asumimos que no hay desviaciones del equilibrio Hardy-Weinberg, no hay desequilibrio de ligamiento, ni drop-outs, ni drop-ins, ni alelos silentes, ni hay ninguna mutación, a menos que se indique explícitamente.

Introducción

El tema del desafío de este año es el ligamiento (*linkage*) y el efecto de los marcadores ligados en las pruebas de parentesco. Para mayor comodidad, recordamos aquí algunos conceptos básicos; para más detalles, se incluyen algunas referencias al final.

La *tasa de recombinación* θ entre dos marcadores es la fracción, a largo plazo, de gametos (productos meióticos) cuyos alelos en los marcadores tienen diferente origen parental. Por ejemplo, la Figura 1 muestra a una madre con los genotipos A/a y B/b en dos marcadores. Para simplificar, asumimos que la fase gamética es conocida, como se indica con la línea vertical. De esta situación pueden surgir cuatro haplotipos diferentes, como se muestra en la parte inferior de la Figura 1 junto con sus probabilidades de transmisión. Los dos haplotipos externos son no recombinantes; los internos son recombinantes.

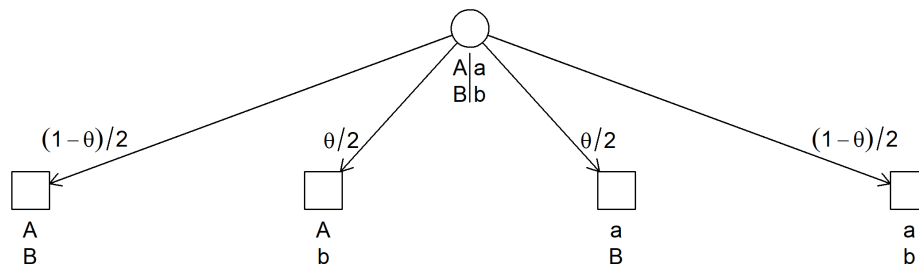


Figure 1. Ilustración de las probabilidades de recombinación y de transmisión de haplotipos en términos de θ .

Para dos loci independientes, por ejemplo, situados en cromosomas diferentes, los cuatro haplotipos de la Figura 1 son igualmente probables, y $\theta = 0.5$. A estos loci se les llama no ligados (*unlinked*). Por el contrario, los loci con $\theta < 0.5$ están ligados (*linked*).

La distancia genética (*genetic map distance*) entre dos loci en el mismo cromosoma se define como el número esperado de entrecruzamientos (*crossovers*) por meiosis entre ellos. La unidad básica es 1 Morgan, que corresponde a 1 entrecruzamiento esperado por meiosis. Una unidad más común es 1 centiMorgan (cM) = 0.01 Morgan.

Existen varios abordajes para convertir tasas de recombinación en distancia genética y viceversa. La más simple es la función de mapeo de Haldane (*Haldane's map function*), que relaciona la tasa de recombinación θ entre dos loci y su distancia genética d medida en Morgans:

$$\theta = \frac{1}{2}(1 - e^{-2d}).$$

Software: A continuación se muestra una lista (no exhaustiva) de programas que admiten marcadores ligados y que pueden ser útiles para este desafío. (Haga click en los enlaces para obtener más información). Se incluyen algunas instrucciones para KLINK al final.

- **KLINK:** Aplicación recientemente desarrollada basada en *pedsuite*, compatible con el software Familias.
- **FamLink2:** Software independiente bien establecido para Windows. Acepta archivos de Familias como entrada.
- **pedsuite:** Conjunto de paquetes de R para análisis de pedigríes, que incluye una amplia gama de funciones para pruebas de parentesco.
- **MERLIN:** Software popular para análisis de ligamiento (*linkage*), aunque no está diseñado específicamente para pruebas de parentesco.

Parte I: Un delicado caso de herencia

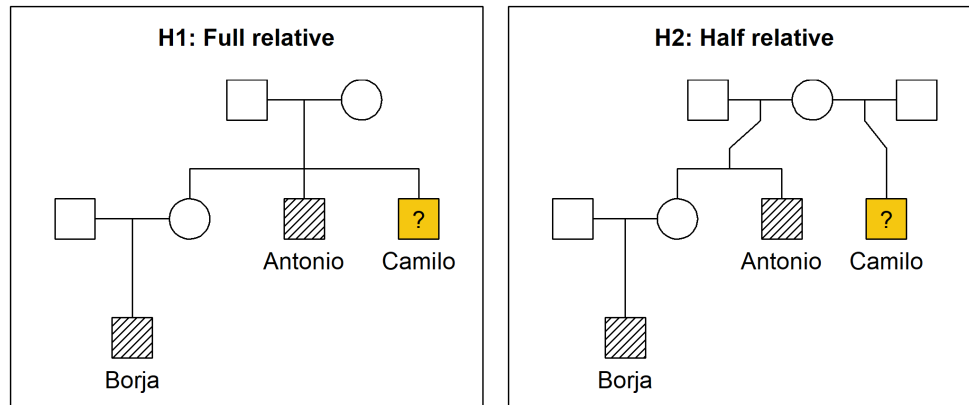
El detective Vargas había visto lo suyo en disputas familiares, pero esta olía a problemas desde lejos. Antonio y su sobrino Borja, ambos con penetrantes ojos azules, se inclinaban sobre su escritorio, discutiendo acaloradamente que Camilo, sentado en silencio junto a ellos, no tenía derecho a la herencia. El padre de Antonio había sido un hombre rico, y ahora había millones en juego.

—Él no es mi hermano completo —afirmó Antonio—. Siempre sospeché que mi madre tenía algo con el jardinero. O con ese malabarista de mala muerte que solía invitar a las fiestas.

—¡Basta con que mires sus ojos marrones! —agregó Borja—. ¡Y ni siquiera puede enrollar la lengua como nosotros!

Ambos comenzaron a enrollar sus lenguas frenéticamente. Vargas cerró los ojos y se recostó en su silla. Los trucos mendelianos para fiestas no resolverían este caso; distinguir entre hermanos completos y medio-hermanos requería un enfoque más científico. ¿Serían suficientes las pruebas estándar de parentesco? Con un suspiro, tomó el teléfono y llamó al mejor genetista forense de la ciudad...

Después de la llamada del detective Vargas, usted esboza los pedigríes que representan las hipótesis alternativas del caso:



Para realizar este ejercicio se dispone de los siguientes archivos:

- *vargasA.ped* y *vargasA.fam*: Conjunto de datos A en formato ped estándar y en formato Familias
- *vargasCombined.ped* y *vargasCombined.fam*: Conjuntos de datos A y B combinados
- *db35.txt*: base de datos de frecuencias para todos los marcadores utilizados en este caso
- *map35.txt*: Mapa indicando el cromosoma y la posición de todos los marcadores en cM

Para un conjunto de datos determinado en los que hay marcadores ligados, introducimos la siguiente notación:

- LR^{ℓ} denota el LR considerando ligamiento, es decir, cuando se tiene en cuenta la dependencia entre los marcadores.

- LR^u denota el LR ignorando el ligamiento, es decir, obtenido simplemente multiplicando los LRs individuales de cada marcador.

Todos los LRs se refieren a las hipótesis H1 yH2 mostradas anteriormente, donde H1 es el numerador y H2 es el denominador.

Inicialmente analiza muestras de los 3 varones con un kit estándar que incluye 19 STRs autosómicos. Este es el conjunto de datos A (Tabla 1).

1. El LR^u total para el conjunto de datos A es aproximadamente:

- 0.71
- 2.01
- 35.78
- 109.23
- 213.07

Table 1: Dataset A

Marker	Antonio	Borja	Camilo
D1S1656	16.3/20.3	16.3/16.3	12/20.3
TPOX	11/11	8/11	8/11
D2S1338	17/23	17/23	17/19
D3S1358	16/16	14/15	16/16
FGA	19/23	20/21	19/20
CSF1PO	10/12	9/12	8/12
D6S474	17/18	14/16	15/16
D7S820	8/12	8/8	10/12
D8S1179	12/13	12/14	12/13
D10S1248	13/16	13/16	14/16
TH01	9/9.3	7/9	9/9.3
vWA	17/18	17/17	17/18
D13S317	12/12	8/12	12/12
PentaE	5/12	5/12	12/20
D16S539	9/11	9/11	9/11
D18S51	19/20	13/16	14/19
D19S433	13/14	14/15	13/14
PentaD	11/12	9/11	11/12
D22S1045	15/16	16/16	11/15

Table 2: Dataset B

Marker	Antonio	Borja	Camilo
D3S3045	11/15	11/14	9/15
D5S2800	17/17	14/17	14/17
D7S3048	21/24	21/23	21/24
D8S1132	19/22	20/22	19/22
D9S1122	12/13	12/12	10/13
D10S1435	11/12	12/13	11/12
D11S2368	19/20	18/19	19/20
D12S391	20/24	17/24	17/20
D13S325	18/19	18/19	18/19
D14S1434	10/11	10/10	10/11
D15S659	17/18	14/18	17/18
D17S1301	11/11	11/12	11/12
D18S1364	13/18	14/14	13/16
D19S253	7/12	12/13	7/12
D20S482	11/14	14/15	12/16
D21S11	28/31.2	28/28	28/30

2. Dos marcadores del conjunto de datos A están ubicados en el mismo cromosoma. Según la función de mapeo de Haldane, la tasa de recombinación entre ellos, redondeada a dos decimales, es:

- 0.41
- 0.42
- 0.45
- 0.48
- 0.49

3. Sea FC el *factor de cambio*¹ del cociente LR^{ℓ}/LR^u para los dos marcadores de la pregunta anterior. Entonces,

- $FC < 1$
- $FC = 1$
- $1 < FC \leq 1.0001$
- $1.0001 < FC \leq 1.001$
- $FC > 1.001$

¹ **N. del T.:** El original en inglés es "Let FC denote the *fold change* ratio LR^{ℓ}/LR^u for the two markers...". El término *fold change* se suele traducir como "**cambio de pliegue**" o "**cambio relativo**", pero en muchos contextos científicos se deja en inglés. El traductor ha decidido utilizar la traducción "factor de cambio" pero también podría usarse "coeficiente de cambio". Básicamente define las veces que es mayor uno de los LRs comparado con el otro. Creemos que el autor ha decidido usar esta definición porque el LR en sí ya es una ratio, y ha considerado necesario aclarar que FC es una ratio de ratios.

Decide analizar a los tres individuos con otro kit, añadiendo 16 nuevos marcadores STR. Este es el conjunto de datos B (Tabla 2). De aquí en adelante, se dice que dos marcadores forman un *par ligado* si están ubicados en el mismo cromosoma.

4. El número de pares ligados en el conjunto de datos total (incluido el par del Ejercicio 2) es:
 - a) 11
 - b) 12
 - c) 13
 - d) 14
 - e) 15
5. El LR^{ℓ} total para los 35 marcadores, después de redondear, es:
 - a) 213
 - b) 1 031
 - c) 7 624
 - d) 15 401
 - e) 23 125
6. El factor de cambio total (FC) del cociente LR^{ℓ}/LR^u teniendo en cuenta los 35 marcadores, es aproximadamente:
 - a) 0.55
 - b) 1.01
 - c) 2.02
 - d) 3.03
 - e) 4.04
7. El número de pares ligados para los cuales LR^{ℓ} es mayor que LR^u , es:
 - a) 4
 - b) 6
 - c) 8
 - d) 10
 - e) 12
8. El número de pares ligados para los cuales LR^{ℓ} y LR^u apoyan hipótesis *diferentes*, es:
 - a) 0
 - b) 1
 - c) 2
 - d) 3
 - e) 4

Una medida adecuada de cuánto afecta el ligamiento al LR es el valor absoluto del logaritmo en base 2 del factor de cambio, es decir:

$$W = \left| \log_2(LR^{\ell}/LR^u) \right|.$$

9. Calcule el valor de W para cada par ligado. Entonces,
 - a) el par más cercano tiene el mayor W y el par más lejano tiene el menor W .
 - b) el par más cercano tiene el mayor W , pero el par más lejano no tiene el menor W .
 - c) el par más cercano no tiene el mayor W , pero el par más lejano tiene el menor W .
 - d) el par más lejano tiene el mayor W .
 - e) ninguna de las anteriores.

En su informe para el detective Vargas, llega a una conclusión basada en el conjunto de datos completo. El protocolo de su laboratorio considera que un $LR > 10\,000$ es un fuerte apoyo para H_1 y un $LR < 0.0001$ es un fuerte apoyo para H_2 ; los umbrales de 1000 y 0.001 proporcionan un apoyo moderado. Los valores entre estos umbrales se consideran inconcluyentes.

Si LR^{ℓ} y LR^u llevan a conclusiones diferentes, debe decidir cómo manejar esta situación.

10. Concluye que los datos:

- a) apoyan fuertemente el parentesco completo (como en H1)
- b) apoyan moderadamente el parentesco completo
- c) apoyan fuertemente el medio parentesco (como en H2)
- d) apoyan moderadamente el medio parentesco.
- e) proporcionan evidencia inconcluyente.

Parte II: Linkage Lab

Los ejercicios siguientes se basan en la aplicación Shiny **Linkage Lab**, disponible on-line en <https://magnusdv.shinyapps.io/linkagelab/>. Esta App está diseñada para explorar el efecto del ligamiento en casos simples de parentesco.

Cada escenario compara dos posibles relaciones entre los individuos A y B, quienes han sido genotipados para dos marcadores, M1 y M2. Ambos marcadores tienen cuatro alelos, etiquetados como 1, 2, 3 y 4, cuyas frecuencias alélicas y tasas de mutación pueden ser establecidas por el usuario (se asume que estos parámetros son los mismos para ambos marcadores). La distancia entre los marcadores puede expresarse en centiMorgans o como tasa de recombinación.

La aplicación calcula la razón de verosimilitud (LR) para las dos relaciones y presenta los resultados en un gráfico, donde uno de los parámetros se puede variar libremente.

Todas las preguntas a continuación se refieren a la comparación hermanos completos vs. medio hermanos (*Sibs* : *half-sibs*). Inicialmente, establezca todas las frecuencias alélicas en 0.25 y la tasa de mutación en 0.

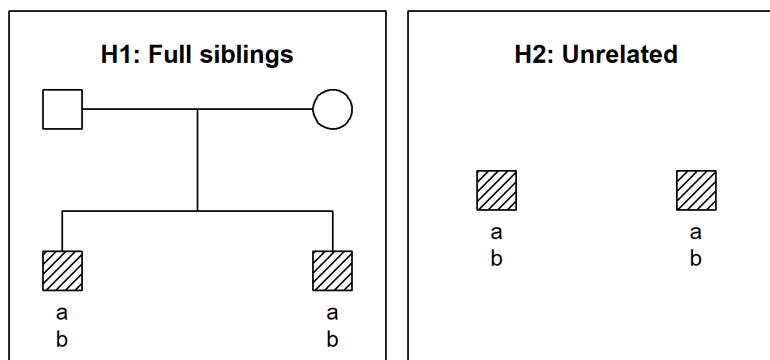
11. Supongamos que A tiene el genotipo 1/2 para ambos marcadores, mientras que B es 3/4 para ambos marcadores. Entonces, *no* es cierto que:
- a) el LR aumenta a medida que el ligamiento es más fuerte
 - b) el LR, considerando ligamiento completo, es exactamente el doble que considerando ausencia de ligamiento
 - c) para ambas hipótesis, la probabilidad es mayor cuando la tasa de recombinación es 0.
 - d) los datos apoyan la hipótesis de hermanos completos si la tasa de recombinación es 0.
 - e) los datos apoyan la hipótesis de medio-hermanos si la tasa de recombinación es 0.5.
12. Ahora cambie el genotipo de B a 1/2 en M1. Entonces, *no* es cierto que:
- a) el LR aumenta a medida que el ligamiento se vuelve más frágil.
 - b) el LR, en ausencia de ligamiento, es exactamente el doble que cuando el ligamiento es completo.
 - c) para ambas hipótesis, la probabilidad es mayor cuando la tasa de recombinación es 0.5.
 - d) los datos apoyan la hipótesis de medio-hermanos si la tasa de recombinación es 0.
 - e) los datos apoyan la hipótesis de hermanos completos si la tasa de recombinación es 0.5.
13. Ahora cambie el genotipo de B a 1/3 para M2 (manteniendo M1 como en el ejercicio anterior). Entonces, *no* es cierto que:
- a) el LR es menor que 1 si los marcadores están completamente ligados.
 - b) el LR es mayor si la tasa de recombinación es 0.25.
 - c) la probabilidad del Ped 1 depende del ligamiento, pero la probabilidad del Ped 2 no.
 - d) los datos apoyan la hipótesis de hermanos completos si los marcadores no están ligados.
 - a) el LR es igual a 1 para marcadores separados a x cM, para un valor determinado de x entre 0 y 10.

Finalmente, supongamos que A y B son ambos 1/1 para M1, y 2/2 y 3/3, respectivamente, para M2. Defina la variable *Plot variable* como "Frequency of '1' allele" (frecuencia del alelo 1), la distancia entre los marcadores en 50 cM y la tasa de mutación en 0.005. Mantenga las frecuencias de los alelos '2' y '3' en 0.25 como antes.

14. La frecuencia de '1' para la cual el $LR = 1$, es aproximadamente
- 0.1
 - 0.2
 - 0.3
 - 0.4
 - 0.5

Parte III: Ligamiento en X

Los ejercicios restantes exploran el efecto del ligamiento entre SNPs del cromosoma X cuando se cuestiona la hermandad. Supongamos que dos niños son hemigóticos para el alelo a en el marcador X1, y para el alelo b en el marcador X2, como se muestra en la figura de abajo.



Sean $p = P(a)$ y $q = P(b)$ las frecuencias alélicas de a y b . En todos los ejercicios a continuación, asumimos que p y q tienen un valor cercano a 0. Sea θ la tasa de recombinación entre los marcadores.

15. La probabilidad de que una mujer aleatoria porte el alelo a en el locus X1 es aproximadamente:
- $p^2 - p$
 - $p/2$
 - p
 - $2p$
 - p^2
16. La probabilidad de la hipótesis H_1 , ignorando un factor constante², se aproxima mejor por:
- $pq\theta(1 - \theta)$
 - $pq(1 - \theta)^2$
 - $p^2q^2(1 - \theta)^2$
 - $pq(\theta^2 + (1 - \theta)^2)$
 - $p^2q^2(\theta^2 + (1 - \theta)^2)$
17. El LR comparando H_1 con H_2 , ignorando un factor constante, se aproxima mejor por:
- $(\theta^2 + (1 - \theta)^2)/(pq)$
 - $1/(pq)$
 - $pq/(\theta^2 + (1 - \theta)^2)$
 - $pq/(\theta(1 - \theta))$
 - pq

² Por ejemplo, si el valor verdadero es $3pq\theta$, entonces $pq\theta$ se considera correcta. (El objetivo de esta formulación es evitar soluciones triviales comprobando las alternativas directamente).

18. En esta situación,

- a) el ligamiento completo es imposible a menos que haya ocurrido una mutación.
- b) el ligamiento completo duplica aproximadamente el LR en comparación con marcadores no ligados.
- c) el LR es mayor si $\theta = 0.25$.
- d) el LR es mayor si los marcadores no están ligados.
- e) el LR es aproximadamente independiente del ligamiento.

Ahora considere la misma situación, pero donde el niño más a la izquierda tiene el alelo común A para el marcador X1.

19. En esta situación,

- a) el ligamiento completo es imposible a menos que haya ocurrido una mutación.
- b) el ligamiento completo duplica aproximadamente el LR en comparación con marcadores no ligados.
- c) el LR es mayor si $\theta = 0.25$.
- d) el LR es mayor si los marcadores no están ligados.
- e) el LR es aproximadamente independiente del ligamiento.

Finalmente, considera la misma situación, pero donde ambos niños tienen el alelo común A para el marcador X1.

20. En esta situación,

- a) el ligamiento completo es imposible a menos que haya ocurrido una mutación.
- b) el ligamiento completo duplica aproximadamente el LR en comparación con marcadores no ligados.
- c) el LR es mayor si $\theta = 0.25$.
- d) el LR es mayor si los marcadores no están ligados.
- e) el LR es aproximadamente independiente del ligamiento.

BUENA SUERTE!

Magnus

Uso de KLINK

Aquí tiene algunas indicaciones sobre la App **KLINK** para pruebas de parentesco con pares de marcadores ligados:

- La App on-line está disponible en <https://magnusdv.shinyapps.io/klink/>.
- Cargue el archivo .fam con el conjunto de datos que desea analizar.
- Establezca *Marker map* como "Custom" y cargue el archivo de mapa.
- Configure *Map function* como "Haldane".
- Es posible que desee ajustar la opción *Ignore Linkage above (cM)*.
- Haga click en *Calculate LR*. Tenga en cuenta que los resultados incluyen LRs asumiendo tanto la presencia como la ausencia de ligamiento.
- Aumente el número de *Decimals* si necesita mayor precisión.

Lecturas adicionales

Los siguientes libros ofrecen más detalles sobre el ligamiento en general y sobre los marcadores ligados en las pruebas de parentesco.

- Thompson, *Statistical Inference from Genetic Data on Pedigrees*. Institute of Mathematical Statistics, 2000.
- Egeland, Kling & Mostad, *Relationship Inference with Familias and R*. Academic Press, 2016.

Responsable del desarrollo del ejercicio:

Magnus Dehli Vigeland

magnusdv@gmail.com

Traducción:

Lourdes Prieto (lourditasmt@gmail.com)

Comité Ejecutivo del GHEP-ISFG

info@ghep-isfg.org