

Desequilibrio de ligamiento

Dr. M. Humberto Reyes - Valdés

■ Asociación

Se dice que ocurre asociación entre dos fenotipos cuando ellos ocurren juntos en el mismo individuo más frecuentemente que lo esperado por azar. La intensidad de esta asociación es frecuentemente medida por la incidencia relativa o riesgo relativo

El riesgo relativo se define como:

$$R = \frac{n_1 (n_2 + n_4)}{n_2 (n_1 + n_3)}$$

donde n_1 es el número de pacientes con una cierta característica, n_2 es el número de individuos control con la característica, y n_3 y n_4 son los números correspondientes sin la característica

Frecuentemente la razón de las "ventajas" (odds ratio) se usa como aproximación:

Razón de las ventajas

$$RO = \frac{n_1 n_4}{n_2 n_3}$$

Otra medida de la intensidad de una asociación es el coeficiente de correlación

Coeficiente de correlación

$$\rho = \frac{N n1 - (n1+n2) (n1+n3)}{\sqrt{(n1+n2) (n1+n3) (n2+n4) (n3+n4)}},$$

Donde $N = n1+n2+n3+n4$

El valor de Ji cuadrada para contingencia es igual a $N \rho^2$

Programas

```
R[n1_, n2_, n3_, n4_] := n1 (n2 + n4) / (n2 (n1 + n3))
```

```
RO[n1_, n2_, n3_, n4_] := n1 n4 / (n2 n3)
```

```
CorrAssoc[n1_, n2_, n3_, n4_] := ((n1 + n2 + n3 + n4) * n1 - (n1 + n2) (n1 + n3)) /  
Sqrt[(n1 + n2) (n3 + n4) (n1 + n3) (n2 + n4)]
```

Ejemplo

Suponga que de 30 pacientes con úlcera gástrica 10 son del grupo sanguíneo A, mientras que de 40 individuos control 11 son del grupo A. Calcule el riesgo relativo como una medida de asociación entre úlcera gástrica y grupo A. Calcule además el odds ratio y la correlación.

```
R[10, 11, 20, 29] // N
```

```
1.21212
```

```
RO[10, 11, 20, 29] // N
```

```
1.31818
```

```
CorrAssoc[10, 11, 20, 29] // N
```

```
0.0629941
```

■ Apareamiento aleatorio con dos loci

Cuando se considera un locus simple, hay dos principios importantes del apareamiento aleatorio:

- a) La probabilidad de un cigoto es el producto de las probabilidades de los dos componentes gaméticos, multiplicados por 2 si la combinación puede ocurrir en dos formas, como lo es el caso de los heterocigotos.
- b) El equilibrio de Hardy-Weinberg se alcanza en una generación después de que inicia el apareamiento aleatorio

Cuando se consideran dos o más loci simultáneamente, el primer principio es válido, mas no el segundo.

En el estado de equilibrio para dos loci, la frecuencia del haplotipo AB es $p_A p_B$, y cuando esto ocurre se dice que la población está en **equilibrio de ligamiento** o, más propiamente, **equilibrio gamético**.

Sean i, j dos alelos situados en diferentes loci. Cuando la frecuencia del haplotipo ij está dada por:

$$P_{ij} = p_i p_j,$$

se dice que la población está en **equilibrio de ligamiento** o **equilibrio gamético** para los loci en cuestión

La razón por la cual el término "equilibrio gamético" es más apropiado, es porque la definición incluye a loci no ligados, i.e., loci en diferentes cromosomas. Cuando la igualdad anterior no se cumple, se dice que existe desequilibrio de ligamiento o desequilibrio gamético para los loci en cuestión.

Causas del desequilibrio de ligamiento

Ligamiento. El fenómeno de ligamiento, acoplado con otros factores, es causante de desequilibrio gamético. La aproximación al equilibrio es mas lenta mientras más ligamiento exista.

Muestreo aglomerado. Si en una investigación los individuos muestreados no son total independientes, e.g. cuando se muestrean varios individuos por familia, se pueden obtener resultados artificiosos de desequilibrio gamético.

Interacción espistática. La interacción entre alelos en diferentes loci puede ocasionar asociaciones entre loci.

Mezcla. Si se mezclan dos poblaciones con diferentes frecuencias haplotípicas, aunque éstas se encuentren en equilibrio gamético, la población resultante está en desequilibrio.

Mutación. Si una mutación que ocurre en un cierto haplotipo, y su frecuencia se incrementa, ocasionará un estado de desequilibrio entre el locus mutante y otros loci de referencia. Obviamente, el ligamiento jugará un papel muy importante en el mantenimiento del estado de desequilibrio.

Aproximación al equilibrio

Sea $P_{AB,t}$ la frecuencia del haplotipo AB en la generación t. Entonces, la siguiente relación existe entre dos generaciones consecutivas:

$$P_{AB,t} = (1 - r) P_{AB,t-1} + p_A p_B,$$

donde r es la frecuencia de recombinación entre dos loci.

Esto nos conduce a la siguiente generalización:

Sean i, j dos alelos situados en diferentes loci. La diferencia entre la frecuencia del haplotipo ij en la generación t y su frecuencia esperada bajo equilibrio gamético está relacionada con dicha diferencia en la generación 0 de la siguiente forma:

$$P_{ij,t} - p_i p_j = (1 - r)^t (P_{ij,0} - p_i p_j)$$

Ya que $P_{ij,t} - p_i p_j$ es una medida del desequilibrio D , entonces

$$D_t = (1 - r)^t D_0$$

Como puede verse, el desequilibrio D , decrece en una magnitud $1-r$ en cada generación de apareamiento aleatorio. Si hacemos $(1 - r)^t = 1/2$, entonces llegamos a la expresión para la vida media del desequilibrio, es decir, el número de generaciones que tarda la población en disminuir a la mitad su desequilibrio original:

$$t = - \frac{\log(2)}{\log(1-r)}$$

Medidas de desequilibrio

Para el caso de dos loci con dos alelos cada uno, las siguientes medidas de desequilibrio de ligamiento son equivalentes

a) $D = P_{AB} - p_A p_B$

b) $D = P_{AB} P_{ab} - P_{Ab} P_{aB}$

El equilibrio es caracterizado por $D = 0$. ¿Cuáles son los límites inferior y superior para D ?

Consideremos las siguientes igualdades:

$$P_{AB} = p_A p_B + D$$

$$D + p_A p_B$$

$$P_{Ab} = p_A p_b - D$$

$$p_A p_b - D$$

$$P_{aB} = p_a p_B - D$$

$$p_a p_B - D$$

$$P_{ab} = p_a p_b + D$$

$$D + p_a p_b$$

Entonces:

$$(p_A p_b - D) \geq 0. \text{ Por lo tanto } D \leq p_A p_b$$

$$(p_a p_B - D) \geq 0. \text{ Por lo tanto } D \leq p_a p_B$$

De aquí que:

$$D \leq \min(p_A p_b, p_a p_B)$$

Ahora:

$$(p_A p_b + D) \geq 0. \text{ Por lo tanto } D \geq -p_A p_b$$

$$(p_a p_B + D) \geq 0. \text{ Por lo tanto } D \geq -p_a p_B$$

De aquí que:

$$D \geq \max(-p_A p_b, -p_a p_B)$$

Con base en los límites superior e inferior de D , Lewontin (1964) propuso una medida estandarizada del desequilibrio:

$$D' = \begin{cases} \frac{D}{\min(p_A p_b, p_a p_B)}, & \text{para } D < 0 \\ \frac{D}{\min(p_A p_b, p_a p_B)}, & \text{para } D \geq 0 \end{cases}$$

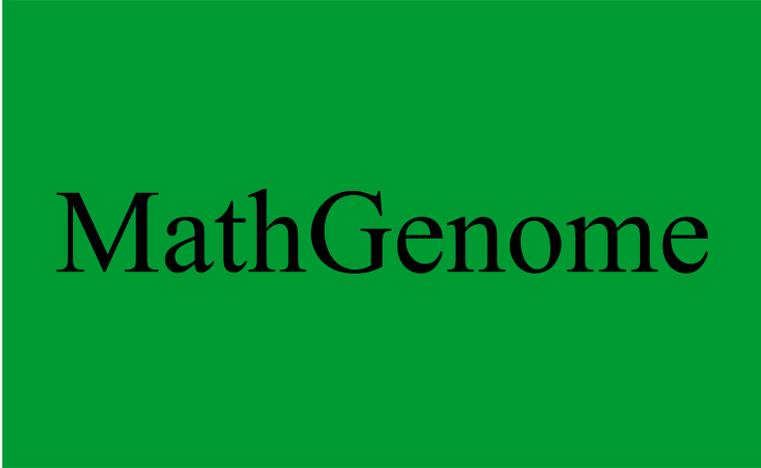
Otra medida de desequilibrio de ligamiento es ρ^2 :

$$\rho^2 = \frac{D^2}{p_A p_B p_a p_b}$$

Esta medida de desequilibrio es idéntica a χ^2/n y se denota frecuentemente por Δ^2 . Otra medida que se utiliza es el nivel de significancia empírico p en la λ^2 para asociación. Sin embargo esta es una función estrictamente monótonica de ρ^2 , y al igual que esta, se espera que tenga un comportamiento subóptimo.

Programas

```
In[1]:= << c:\dres\math\mathgenome.m
```



The image shows a green rectangular box with the word "MathGenome" written in a large, black, serif font in the center.

This Mathematica application is being developed by

Humberto Reyes-Valdés

All rights reserved

```
In[2]:= GamDesT[r_, de_, t_] := (1 - r) ^ t de
```

```
In[4]:= GamDes[AB_, Ab_, aB_, ab_] := AB ab - Ab aB
```

```
In[6]:= GamDesSt[AB_, Ab_, aB_, ab_] :=
  (pofA = AB + Ab; pofB = AB + aB; If[GamDes[AB, Ab, aB, ab] < 0,
    GamDes[AB, Ab, aB, ab] / Min[pofA pofB, (1 - pofA) (1 - pofB)],
    GamDes[AB, Ab, aB, ab] / Min[pofA (1 - pofB), (1 - pofA) pofB])
```

```
In[7]:= GamDesHalf[r_] := IntegerPart[-Log[2] / Log[1 - r]] + 1
```

```
In[8]:= GamDesRoS[AB_, Ab_, aB_, ab_] := (pofA = AB + Ab; pofB = AB + aB;
  GamDes[AB, Ab, aB, ab] ^ 2 / (pofA (1 - pofA) pofB (1 - pofB)))
```

```
In[10]:= GamDesP[AB_, Ab_, aB_, ab_, n_] :=
  1 - CDF[ChiSquareDistribution[1], GamDesRoS[AB, Ab, aB, ab] n] // N
```

Ejemplos

Si se tiene las siguientes frecuencias de haplotipos: AB (.3), Ab (.2), aB(.1), ab(.4). Calcule a) D , b) D' , c) el número de generaciones para que el desequilibrio se reduzca a la mitad, asumiendo $r = .1$, d) la magnitud del desequilibrio en la generación 10 de apareamiento aleatorio, e) ρ^2 y f) p

```
In[11]:= GamDes [.3, .2, .1, .4]
```

```
Out[11]= 0.1
```

```
In[12]:= GamDesSt [.3, .2, .1, .4]
```

```
Out[12]= 0.5
```

```
In[13]:= GamDesHalf [.1]
```

```
Out[13]= 7
```

```
In[14]:= GamDesT [.1, .1, 10]
```

```
Out[14]= 0.0348678
```

Note que PAB-pA pB da el mismo resultado para D . Por otro lado, si se invierten las frecuencias de haplotipos, tenemos:

```
In[15]:= GamDes [.2, .3, .4, .1]
```

```
Out[15]= -0.1
```

```
In[16]:= GamDesSt [.2, .3, .4, .1]
```

```
Out[16]= -0.5
```

```
In[17]:= GamDesRoS [.2, .3, .4, .1]
```

```
Out[17]= 0.166667
```

```
In[18]:= GamDesP [.2, .3, .4, .1, 100]
```

```
Out[18]= 0.0000445571
```

Funciones de mapeo

Nota: Material tomado de Reyes-Valdés MH, 2003. Mapping functions. In: Manjit S. Kang, Editor. Handbook of formulas and software for geneticists and breeders. The Haworth Reference Press. New York. p. 255-260.

■ Definiciones

Funcion de mapeo. Una función matemática que relaciona la distancia mapa con la frecuencia de recombinación.

Distancia genética mapa (en morgans). El número promedio de intercambios por evento meiótico entre dos loci. La distancia genética se relaciona con la distancia física, pero ambas no son equivalentes.

Unidad Morgan. Unidad para expresar la distancia entre loci genéticos, basada en recombinación. El término fue acuñado por Haldane (1919) en honor a T. H. Morgan.

Coincidencia. Dobles recombinaciones observadas/número esperado si la distribución es aleatoria. Cada función de mapeo tiene una suposición explícita o implícita acerca de la coincidencia.

■ Funciones

Nota: m = distancia mapa; y = frecuencia de recombinación

Haldane (1919). Asume $C = 1$

$$m = - (1 / 2) \text{Log}[1 - 2 y]$$

$$-\frac{1}{2} \log(1 - 2 y)$$

Kosambi (1944). Asume $C = 2y$

$$m = (1 / 2) \text{ArcTanh}[2 y]$$

$$\frac{1}{2} \tanh^{-1}(2 y)$$

Pascoe y Morton (1987). Asume $C = (2y)^2$

$$m = (-\text{Log}[(1 - 2 y)^2 / (1 + 2 y + 4 y^2)] / 12) +$$

$$(\text{Sqrt}[3] \text{ArcTanh}[(1 + 4 y) / \text{Sqrt}[3]] / 6) - 0.15115$$

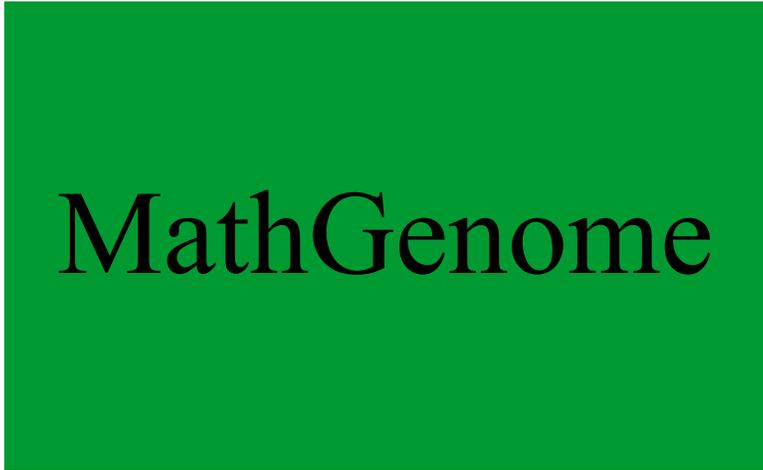
$$\frac{\tanh^{-1}\left(\frac{4y+1}{\sqrt{3}}\right)}{2\sqrt{3}} - \frac{1}{12} \log\left(\frac{(1-2y)^2}{4y^2+2y+1}\right) - 0.15115$$

Carter and Falconer (1951). Asume $C = (2y)^3$

$$m = (1 / 4) [\text{ArcTanh}[2 y] + \text{ArcTan}[2 y]]$$

$$\frac{1}{4} [\tan^{-1}(2 y) + \tanh^{-1}(2 y)]$$


```
<< c:\dres\math\mathgenome.m
```



This Mathematica application is being developed by

Humberto Reyes-Valdés

All rights reserved

```
Map[HFun, {0.05, 0.10, 0.15, 0.20}] 100
{5.26803, 11.1572, 17.8337, 25.5413}

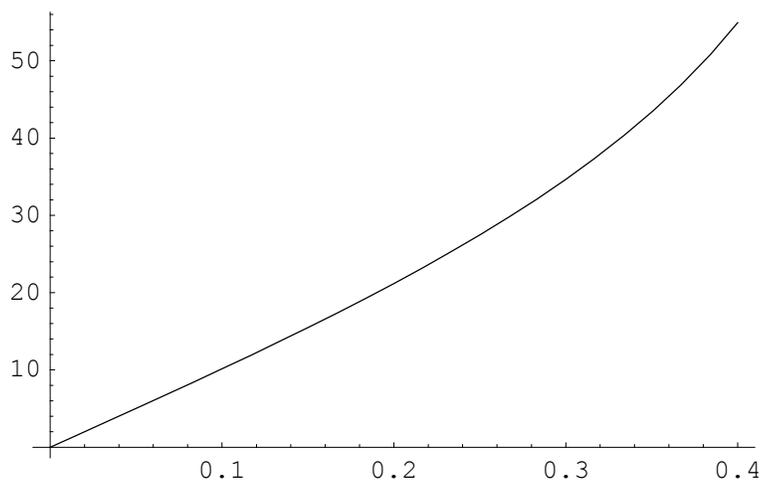
{Map[HFun, {0.05, 0.10, 0.15, 0.20}], Map[KFun, {0.05, 0.10, 0.15, 0.20}],
 Map[PMFun, {0.05, 0.10, 0.15, 0.20}],
 Map[CFun, {0.05, 0.10, 0.15, 0.20}]} 100
( 5.26803 11.1572 17.8337 25.5413 )
( 5.01677 10.1366 15.476 21.1824 )
( 5.00125 10.0201 15.1028 20.3322 )
( 5.0001 10.0032 15.0244 20.1039 )
```

2. Convertir las siguientes distancia mapa en cM a frecuencia de recombinación (10, 20, 30)

```
{Map[IHFun, {10, 20, 30} / 100], Map[IKFun, {10, 20, 30} / 100],
 Map[IPMFun, {10, 20, 30} / 100],
 Map[ICFun, {10, 20, 30} / 100]} // N
( 0.0906346 0.16484 0.225594 )
( 0.0986877 0.189974 0.268525 )
( 0.0998006 0.196886 0.285161 )
( 0.0999684 0.198987 0.292644 )
```

3. Hacer una gráfica de la función de Kosambi, para recombinaciones de 0 a 0.4. Mostrar en el eje de las y la distancia mapa en cM.

```
Plot[100 KFun[x], {x, 0, 0.4}]
```



- Graphics -