Probabilidad de obtener segregaciones esperadas

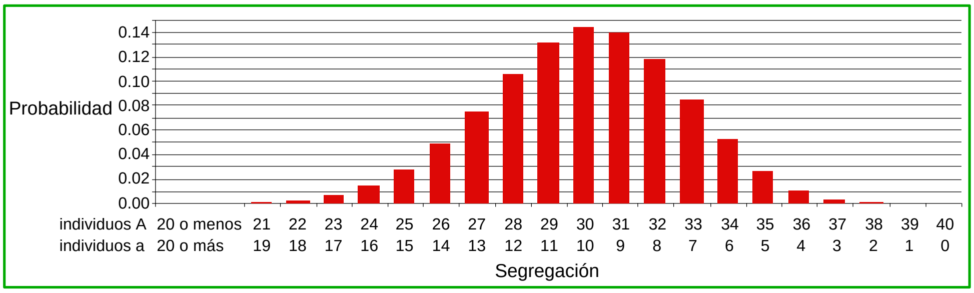
El análisis de segregaciones es un tema crucial en Genética que incluye la realización de observaciones y la comparación de los datos obtenidos con segregaciones esperadas bajo diferentes supuestos teóricos.

Por ejemplo, en la descendencia del cruzamiento entre dos heterozigotos para un gen con dominancia completa (Aa x Aa) la probabilidad teórica de aparición de un individuo con fenotipo dominante es 3/4 y la de un individuo con fenotipo recesivo es 1/4. Si la descendencia está compuesta por 40 individuos, la segregación esperada con la que debe compararse la observada a la hora de hacer un test χ^2 es: 30A:10 a.

Sin embargo, afirmar que 30A:10a es la distribución esperada, es una simplificación del problema. En realidad, lo que se espera en este caso es lo que figura en la tabla de la derecha. Se trata de 40 individuos, y cada uno de ellos puede presentar las dos alternativas A o a con probabilidades 3/4 y 1/4, respectivamente. Por tanto, se espera que la segregación observada sea una de las 41 clases (segregaciones) diferentes que aparecen en la tabla, cuyas probabilidades son las de la correspondiente distribución binomial :

$$f_i = \frac{40!}{i! \times (40-i)!} \times \left(\frac{3}{4}\right)^i \times \left(\frac{1}{4}\right)^{40-i}$$

En la siguiente figura se presentan los datos de la tabla en forma gráfica.



Clase (i)	Segregación iA : (40-i)a	Probabilidad (f_i)
0	0A : 40a	f_0= 8.27x10^{-25}
1	1A : 39a	f_1= 9.92x10^{-23}
2	2A : 38a	f_2= 5.81x10^{-21}
3	3A : 37a	f_3= 2.21x10^{-19}
4	4A : 36a	f_4= 6.12x10^{-18}
5	5A : 35a	f_5= 1.32x10^{-16}
6	6A : 34a	f_6= 2.31x10^{-15}
7	7A : 33a	f_7= 3.37x10^{-14}
8	8A : 32a	f_8= 4.17x10^{-13}
9	9A : 31a	f_9= 4.45x10^{-12}
10	10A : 30a	f_{10}= 4.14x10^{-11}
11	11A : 29a	f_{11}= 3.39x10^{-10}
12	12A : 28a	f_{12}= 2.46x10^{-9}
13	13A : 27a	f_{13}= 1.59x10^{-8}
14	14A : 26a	f_{14}= 9.18x10^{-8}
15	15A : 25a	f_{15}= 4.77x10^{-7}
16	16A : 24a	f_{16}= 2.24x10^{-6}
17	17A : 23a	f_{17}= 9.48x10^{-6}
18	18A : 22a	f_{18}= 3.63x10^{-5}
19	19A : 21a	f_{19}= 0.000126
20	20A : 20a	f_{20}= 0.000397
21	21A : 19a	f_{21}= 0.001136
22	22A : 18a	f_{22}= 0.002943
23	23A : 17a	f_{23}= 0.006910
24	24A : 16a	f_{24}= 0.014684
25	25A : 15a	f_{25}= 0.028192
26	26A : 14a	f_{26}= 0.048794
27	27A : 13a	f_{27}= 0.075903
28	28A : 12a	f_{28}= 0.105721
29	29A : 11a	f_{29}= 0.131240
30	30A : 10a	f_{30}= 0.144364
31	31A : 9a	f_{31}= 0.139707
32	32A : 8a	f_{32}= 0.117878
33	33A : 7a	f_{33}= 0.085730
34	34A : 6a	f_{34}= 0.052951
35	35A : 5a	f_{35}= 0.027232
36	36A : 4a	f_{36}= 0.011347
37	37A : 3a	f_{37}= 0.003680
38	38A : 2a	f_{38}= 0.000872
39	39A : 1a	f_{39}= 0.000134
40	40A : 0a	f_{40}= 1.01x10^{-5}

Es interesante destacar que, aunque la clase 30 (segregación 30A:10a) tiene la probabilidad máxima, esa probabilidad es sólo f_{30}= 0.144. Es decir, si se analizaran muchas segregaciones de este tipo, obtendríamos la segregación 30A:10a menos de 15 veces de cada cien. De hecho, sería mucho mas probable obtener cualquier otro resultado.

Por tanto, si es posible cualquier resultado, ¿qué criterio debe usarse para establecer hasta qué punto una segregación observada cumple o nó una hipótesis?

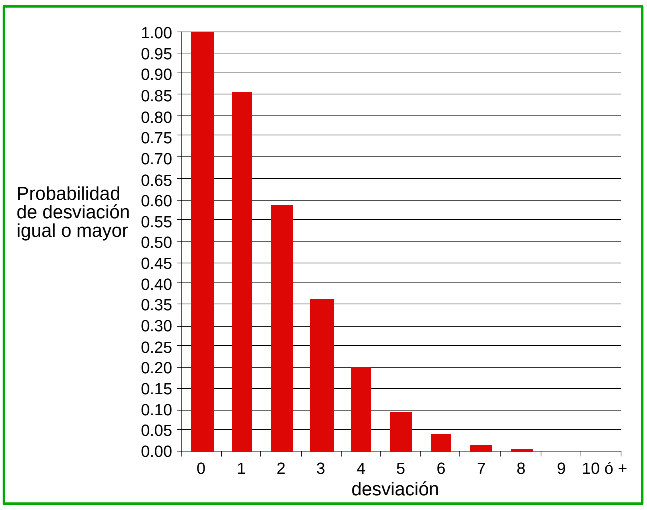
El problema puede tratarse considerando la desviación de la segregación observada frente a la segregación más frecuente y la probabilidad de obtener una desviación igual o mayor que la observada. Por ejemplo, la segregación 30A:10a es la clase más frecuente (se desvía 0 unidades de ella misma), por lo que cualquier otra segregación mostrará mayor desviación y, por tanto la probabilidad de obtener una desviación igual o mayor que 0 será 1. Las segregaciones 29A:11a y 31A:9a se desvían una unidad de la clase más frecuente (30A:10a), y la probabilidad de obtener una desviación igual o mayor que 1 es: 1-f_{30}= 0.856. Las segregaciones 28A:12a y 32A:8a se desvían dos unidades de la clase más frecuente (30A:10a), y la probabilidad de obtener una desviación igual o mayor que 2 es: 1 - f_{30} - f_{29} - f_{31}= 0.585. En la gráfica inferior derecha se indican las probabilidades de obtener segregaciones que muestren desviaciones iguales o mayores a 0, 1, 2, 3, etc.

Ahora, podríamos acordar que la desviación observada (el resultado obtenido) se ajusta a la hipótesis, si la probabilidad de obtener una desviación igual o mayor que la obtenida es razonablemente alta.

Por convenio, se considera que la desviación es significativa (los datos observados no se ajustan a los esperados) si la probabilidad de obtener una desviación igual o mayor que la observada es inferior a 0.05.

Esa probabilidad 0.05 significa que si se repitiera el experimento muchas veces y la hipótesis fuera cierta, sólo cinco de cada cien veces se obtendría una desviación igual o mayor que la observada. Si el experimento se hace sólo una vez y los datos observados son muy poco probables en relación con la hipótesis, lo más razonable es rechazar esa hipótesis. Se considera que la desviación es altamente significativa (el ajuste es aún peor) si la probabilidad de obtener una desviación igual o mayor que la observada es inferior a 0.01.

Imaginemos ahora que, en el ejemplo que nos ocupa (descendencia de 40 individuos con segregación esperada 3:1), se hace el experimento y se obtiene una segregación 25A:15a. Esa segregación se desvia 5 unidades frente a la segregación 30A:10a y en la gráfica de la derecha puede comprobarse que la probabilidad de obtener una desviación igual o mayor que 5 es superior a 0.05. Por tanto, concluiremos que la desviación no es significativa. Sin embargo, si se obtiene una segregación 24A:16a (desviación de 6 unidades frente a la segregación 30A:10a) concluiremos que la desviación es significativa, ya que, tal como puede comprobarse en la gráfica de la derecha, la probabilidad de obtener una desviación igual o mayor que 6 es inferior a 0.05.



Este tipo de cálculos es complicado y lo es aún más cuando se consideran segregaciones con más de dos clases. Sin embargo, es muy fácil utilizar el test χ^2, que supone una aproximación bastante buena al problema:

Efectivamente, al aplicar el test χ^2 a las segregaciones 25A:15a y 24A:16a se obtienen las mismas conclusiones anteriores. Con la hipótesis 3:1, los valores de χ^2 para las segregaciones 25A:15a y 24A:16a son:

$$\chi^2_{3:1} = \frac{(25-30)^2}{30} + \frac{(15-10)^2}{10} = 3.333 \qquad \chi^2_{3:1} = \frac{(24-30)^2}{30} + \frac{(16-10)^2}{10} = 4.800$$

Para un grado de libertad (véase el test χ^2) el valor $\chi^2_{3:1}$= 3.333 se corresponde con una probabilidad (p) de obtener una desviación igual o mayor que la encontrada de 0.10>p>0.05 (desviación no significativa); y el valor $\chi^2_{3:1}$= 4.800 se corresponde con una probabilidad 0.05>p>0.01 (desviación significativa).