El test χ^2

En numerosos experimentos en Genética, como en los juegos de azar, se obtienen resultados en forma de distribución con clases discontinuas. Por ejemplo, si se tiran N veces tres monedas, aparecerán las siguientes cuatro clases: 3 caras y 0 cruces (clase 0) ; 2 caras y 1 cruz (case 1); 1 cara y 2 cruces (clase 2); 0 caras y 3 cruces (clase 3) con frecuencias O_0, O_1, O_2 y O_3 respectivamente ($O_0 + O_1 + O_2 + O_3 = N$). Del mismo modo, en una F2 de un cruzamiento entre dos homocigotos AA x aa, aparecerán las tres clases genotípicas AA, Aa y aa con frecuencias O_1, O_2 y O_3, respectivamente ($O_1 + O_2 + O_3 = N$).

El test χ^2 (chi cuadrado o ji cuadrado) se usa para establecer hasta qué punto los resultados experimentales (distribución observada) se desvían de resultados esperados (distribución esperada) que se derivan de determinada hipótesis (hipótesis nula).

En el ejemplo de las monedas, si éstas no están trucadas y se comportan de forma independiente, las frecuencias esperadas de las distintas clases (E_0, E_1, E_2 y E_3) serán las de la correspondiente distribución binomial :

$$E_i = \binom{3}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{3-i} N$$

En el caso de la F2 las frecuencias absolutas esperadas serán: $E_1 = \frac{1}{4}N$ $\quad E_2 = \frac{1}{2}N$ $\quad E_3 = \frac{1}{4}N$

El estadístico χ^2 se calcula siempre a partir de frecuencias absolutas (es decir, número de sucesos), nunca a partir de frecuencias relativas (porcentajes o probabilidades). De hecho, parte de la utilidad del test χ^2 es que toma en consideración el tamaño de la muestra. En cualquier caso, la fórmula para calcular el χ^2 es:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \text{ ; es decir, la suma, para todas las clases, de: } \frac{(\text{valor observado - valor esperado})^2}{\text{valor esperado}}$$

Con el valor de χ^2 obtenido, podemos estimar la probabilidad (*p*) de obtener los resultados observados si la hipótesis nula es correcta (véase probabilidad de obtener segregaciones esperadas).

Primero, es necesario conocer el número de grados de libertad (gl). Como regla general, el número de grados de libertad es igual al número de clases de la distribución menos el número de parámetros que se toman de la muestra para construir la correspondiente distribución esperada.

En casos como los anteriormente citados, para elaborar la distribución de valores esperados se toma un sólo parámetro de la muestra (el número total de observaciones, N), por lo que el número de grados de libertad es: gl= número de clases - 1 (gl= 3 en el caso de las monedas; gl=2 en el caso de la F2).

Ahora, debe usarse la tabla de la distribución χ^2. En la fila correspondiente a los grados de libertad del caso que nos ocupa se busca el valor obtenido de χ^2, que se corresponderá con un valor de probabilidad *p* (fila superior). En realidad, p es la probabilidad de obtener una desviación entre los datos observados y esperados igual o superior a la obtenida en el supuesto de que la hipótesis nula fuera cierta y esa desviación se debiera sólo al azar. Por convenio, si *p* es superior a 0.05 se considera que la desviación no es significativa, es decir, que los valores observados se ajustan razonablemente a los esperados bajo la hipótesis nula, por lo que ésta puede aceptarse. Si *p* esta comprendida entre 0.05 y 0.01 se considera que la desviación es significativa, es decir, los resultados se desvían mucho y la hipótesis debe rechazarse. Para valores de *p* inferiores a 0.01 se considera que la desviación es altamente significativa, es decir, que la probabilidad de que la hipótesis sea cierta es bajísima y debe rechazarse con más motivo.

Distribución χ^2

	Probabilidad (*p*)										
gl	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.00	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32

Ejemplo

Se realiza un cruzamiento entre dos líneas puras de plantas con flores color rojo y color blanco, respectivamente. La F1 tiene flores de color rojo y en la F2 se obtiene una segregación compuesta por 115 individuos de flores rojas y 13 de flores blancas. Se quiere saber hasta qué punto esa segregación se corresponde con la esperada para un solo locus (3:1) o para dos loci (15:1).

- El número total de individuos observados es N= 115 + 13= 128
- Los valores esperados en la hipótesis 3:1 para las dos clases son: $\frac{3}{4}128 = 96$ individuos de flores rojas; $\frac{1}{4}128 = 32$ individuos de flores blancas

- El valor del correspondiente χ^2 (1gl) es:

$$\chi^2_{3:1} = \frac{(115 - 96)^2}{96} + \frac{(13 - 32)^2}{322} = 15.04$$

gl	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001	0.001>p
1	0.00	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83	15.04

- En la fila de valores de la tabla para un grado de libertad, puede verse que este χ^2 es mayor que 10.83, por lo que se corresponde con una probabilidad inferior a 0.001 ($p<0.001$). Es decir, los datos obtenidos difieren de forma altamente significativa de los esperados en el supuesto de una segregación 3:1

- Los valores esperados en la hipótesis 15:1 son: $\frac{15}{16}128 = 120$ individuos de flores rojas; $\frac{1}{16}128 = 8$ individuos de flores blancas

- El valor del correspondiente χ^2 (1gl) es:

$$\chi^2_{15:1} = \frac{(115 - 120)^2}{120} + \frac{(13 - 8)^2}{8} = 3.33$$

gl	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.10>p>0.05	0.05	0.01	0.001
1	0.00	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.33	3.84	6.64	10.83

- En la fila de valores de la tabla para un grado de libertad, puede verse que este χ^2 se sitúa entre 2.71 y 3.84, por lo que se corresponde con una probabilidad superior a 0.05 ($0.10>p>0.05$). Es decir, los datos obtenidos no difieren significativamente de los esperados en el supuesto de una segregación 15:1