#007★☆☆☆ 2021.5.25 / 2023.7.23

分割表とカイ二乗検定。

	結果（事象発現など）の有無		
	有	無	合計
要因 X	A	B	A+B
要因 B	C	D	C+D
合計	A+C	B+D	A+B+C+D

公害の多い都市部在住の人たちと比較的空気の澄んだ公害在住の人たちとで肺がん発生率に違いはあるか、同効の薬剤 A と B をそれぞれ投与した患者で副作用発症率に違いはあるか、といった「因果関係」の解析をするとき、カルテ情報などから「分割表」を作成することがある。要因（例えば薬剤の種類）、結果（例えば副作用の有無）がそれぞれ 2 通りのとき 2×2 分割表という。分割表では A〜D の各条件（位置を「セル」と呼ぶ）で実際に測定された人数を「度数」として設定する。例えば次の表では「Drug X 投与後に副作用が発症した患者は 250 人」となる。これらの事象が独立に起こることを前提に、カイ二乗検定、オッズ比、その他の統計指標が得られる。

観測度数	副作用の有無		
	有	無	合計
Drug X	250	75	325
Drug Y	155	120	275
合計	405	195	600

期待度数	副作用の有無		
	有	無	合計
Drug X	219.375	105.625	325
Drug Y	185.625	89.375	275
合計	405	195	600

カイ二乗値	副作用の有無		
	有	無	合計
Drug X	4.28	8.88	
Drug Y	5.05	10.49	
合計			28.70

カイ二乗検定（Chi-squared test）では、観測度数（O）と期待度数（E）から各セルについてカイ二乗値を(O - E)^2 ÷ E で計算し（^2 は二乗の意味）、全てのセル（2×2 分割表の場合 4 つのセル）のカイ二乗値の和を求め（表では 28.70)、この値を表全体の「偏り」の和と考える。期待度数は「もし分割表に偏りがなかったら（=帰無仮説の場合)」の条件での観測数の計算値であり、例えば A のセルの場合、Drug X を投与した合計 325 名が合計の副作用有無の比率で分配された場合として 325×(405/600) = 219.375 と求められる。次に、有意水準（α）を定める。通常の検定では $\alpha=0.05$ または $\alpha=0.01$ とする場合が多い。さらに自由度（df、degree of freedom）は、2×2 分割表においては（「要因」の項目数 2 から 1 を引く）×（「結果」の項目数 2 から 1 を引く）の計算で df =1（= (2 – 1) × (2 – 1)）と得られる。以上の数値を用いて次に示す表と比較して、カイ二乗計算値＞表の（指定された α 及び df の）数値、であれば「有意な因果関係あり」と結論する。今回の例ではカイ二乗計算値（28.70）＞（自由度 1、有意水準 0.05 のときの表の値（3.841）なので「$p<0.05$ で有意な因果関係あり」となる。有意水準を 0.01 とした場合でもカイ二乗計算値のほうが大きいので「$p<0.01$ で有意な因果関係あり」とも言える。

df	1	2	3	4	5
α = 0.05	3.841	5.991	7.815	9.488	11.071
α = 0.01	6.635	9.21	11.345	13.277	15.086

なお、データ数が少ない場合（ひとつの目安として、期待度数が 5 以下のセルがある場合）は Fisher の正確検定（Fisher’s exact test）を利用するべきとされている。