

(a)

	実データ		仮想データ1		仮想データ2	
	罹患群	対照群	罹患群	対照群	罹患群	対照群
①飲まない	222	228	222	228	202	228
②少し飲む	86	105	106	105	86	105
③多く飲む	149	124	129	124	169	124
計	457	457	457	457	457	457
p-value	0.1189		0.9123		0.0056	
-log(p-value)	0.9248		0.0399		2.2526	

(b)

遺伝子型	罹患群	対照群
GG	287	292
GA	149	150
AA	21	15
計	457	457
p-value	0.5926	
-log(p-value)	0.2272	

(c)

```

> data <- [1,] [1,2]
[1,] 222 228
[2,] 86 105
[3,] 149 124
> chisq.test(data)$p.value
[1] 0.1188712
> -log10(0.1188712)
[1] 0.9249234

```

図2.4 コホート研究によって得られたデータの解析例

(a) Hidaka論文中の「実データ」（左図）。原著論文のTable 1中にあるAlcohol consumptionの情報に相当する。右図の「仮想データ1」は疾患群中の灰色部分の数値を対照群中の数値に似せたもの、そして「仮想データ2」は「実データ」よりもさらに対照群とは異なる数値分布にしたものである。一番下は、カイ二乗検定結果として得られたp値を示している。3つのシナリオを眺めることで、対照群の数値分布と似ているほどp値が1に近づき、離れるほど0に近づくことが分かる。(b) ジェノタイピングによって得られた遺伝子型ごとの数値分布。全体としてp値は高めにしていることがわかる。遺伝子型AAを持っている絶対数が少ないものの、持っていない場合に比べて21/15 = 1.4倍胃がんに罹患しやすいことが分かる。(c) データ解析環境RStudioでの実行例。