

## Hard-Weinberg 平衡检验

英国数学家哈代 (G. H. Hardy) 和德国生理学家温伯格 (W. Weinberg) 于 1908 年同时发表了遗传学中的平衡定律, 他们证明了基因型频率在一代随机交配后达到平衡, 而且以后一直保持这种平衡, 除非一些因素 (突变, 迁移和自然选择) 改变群体的等位基因频率, 亦即在一个大的、随机婚配的群体中, 如果没有迁移、选择的影响, 突变率保持不变, 各种基因型的频率也将代代保持不变, 成为遗传平衡的群体, 此即遗传平衡定律, 亦称为 Hardy-Weinberg 定律。

以二倍体的个体组成的群体, 在完全随机交配的前提下, 假设不存在突变、选择和迁移, 基因频率逐代不变, 随机交配一代以后基因型频率将保持平衡, 基因型频率之和应等于 1, 亦即:  $P_A^2 + 2P_AP_B + P_B^2 = 1$ , 其中、 $P_A$  表示 A 基因的频率,  $P_B$  表示 B 基因的频率 ( $P_A + P_B = 1$ ), AA 基因型的频率为  $P_A^2$ , AB 基因型的频率为  $2P_AP_B$ , BB 基因型的频率为  $P_B^2$ 。

Hardy-Weinberg 定律是群体遗传学中的最为重要的原理, 在以群体为研究对象进行单基因遗传病分析时可以用遗传平衡定律来估计群体中所研究疾病的基因频率和基因型频率。在对遗传基因型数据进行分析之前, 一般均需先检测我们采集的数据是否符合 Hard-Weinberg 平衡定律, 以保证数据是随机抽取并具有一定代表性的, 检验方法一般使用拟合优度检验。

**例 1**, 某研究选取血管紧张素原 (angiotensinogen, AGT) 基因启动子-217, -152, -20, -6, 内含子 1 的+31, 共 5 个位点, 对 100 例正常人进行了 SNP 位点基因型检测, (下载数据: [http://www.empowerstats.com/empowerStats/exdata/hardyweinberg\\_data.xls](http://www.empowerstats.com/empowerStats/exdata/hardyweinberg_data.xls)), 其中 5 个位点的基因型分别用 0, 1, 2 代表: G217-A 为 AA, AG 和 GG; G152-A 为 AA, AG 和 GG, A-20C 为 AA, AC 和 CC, A-6G 为 AA, AG 和 GG, C+31T 为 CC, CT 和 TT, 试检验这几个位点基因型频率是否满足 Hardy-Weinberg 定律。

输入界面:

### Hardy-Weinberg平衡检验 ?

标题:

选择分析对象:

基因型/等位基因变量	基因型/等位基因变量类型
变量	1:基因型(每个变量代表一个基因)
G217-A	
G152-A	
A-20C	
A-6G	
C+31T	

选择分层变量:

如果原始数据中是基因型变量，则应该在“基因型/等位基因变量类型”下选择“1:基因型(每个变量代表一个基因)”；如果是等位基因变量，则应该选择“2:等位基因(按顺序2个变量代表一个基因)”。

**输出结果:**

Hardy-Weinberg 平衡检验

基因型变量	#AA	#AB	#BB	P 值	卡方值	D	等位基因 1 频率	等位基因
G217. A	76	22	2	0.88083	0.02247	-0.31000	0.87000	1
G152. A	91	8	1	0.55076	0.35596	-0.75000	0.95000	1
A. 20C	74	24	2	0.73105	0.11815	-0.04000	0.86000	1
A. 6G	61	33	6	0.75766	0.09521	-0.93750	0.77500	1
C. 31T	62	31	7	0.38047	0.76920	-1.93750	0.77500	1

输出的表格中列出了每个基因型变量的详细情况及 Hardy-Weinberg 定律检验的结果，以 G217-A 为例：其中 AA 型：76 例，AG 型：22 例，GG 型：2 例，检验的 P 值：0.88083，卡方值：0.02247，D 值（AB 基因型偏离 Hardy-Weinberg 平衡的距离的 1/2）：-0.31000，等位基因 A 的频率：0.87000，P 值>0.05，基因型频率符合 Hardy-Weinberg 定律。