

分组化验模型的优化研究

高焕江¹, 张翠丽¹, 高 鹏²

(1. 邢台医学高等专科学校数学教研室, 河北 邢台 054000)

(2. 第三军医大学预防医学院, 重庆 400038)

摘 要: 本文研究了血液的分组化验模型. 利用概率和微分学基本理论证明了当血样呈阳性的概率在某一确定范围时, 分组化验存在最佳分组方案, 针对给定的血样阳性概率给出了最佳分组 k 值的确定方法, 得出了分组化验最佳分组 k 值计算公式.

关键词: 分组化验; 数学期望; 最佳分组 k 值; 迭代公式; 阶梯函数

MR(2010) 主题分类号: 60H30

中图分类号: O211.9

文献标识码: A

文章编号: 0255-7797(2014)03-0583-06

1 引言

血样数量很大的化验检查 (如人群疾病普查等) 可采取分组化验, 恰当地分组可减少化验次数降低成本^[1,6]. 自上个世纪八十年代严士健教授在其著述《概率与数理统计基础》一书中提出这个问题以来, 不少数学工作者对此进行探索并取得了一些研究成果^[1-6]. 由于分组化验模型涉及到解超越方程, 而超越方程一般无解析解, 致使目前对这个问题的研究所得出的结论还不够完备, 主要表现在尚未准确得出最佳分组 k 值计算公式. 所谓最佳分组 k 值, 指的是在所有分组方案中使得化验次数最少的方案所对应的每个小组的血样份数. 本文所要解决的问题是, 以相关文献关于分组化验问题的研究结论为基础, 探讨最佳分组 k 值的确定方法, 得出最佳分组 k 值计算公式.

2 分组化验模型

设需化验的血样总份数为 N (N 充分大), 把这 N 份血样分为若干组, 每组 k 份. 将每一组中的 k 份血样各取出一部分混合在一起进行化验: 若混合液呈阴性, 则说明这 k 份血样都呈阴性, 这时这 k 份血样就只需化验一次; 若混合液呈阳性, 则说明这 k 份血样中至少有一份呈阳性, 这时再对这个组的 k 份血样逐份化验, 这种情况下需化验 $k+1$ 次^[1,6]. 这里假定该化验是定性检验, 并且血液混合起来没有交互作用^[4,6].

设血样呈阳性的概率 (先验概率) 为 p ($0 < p < 1$), 呈阴性的概率为 q ($q = 1 - p$), 记每份血样的化验次数为 ξ , 则 ξ 是一个随机变量, 其数学期望为 $E(\xi) = \frac{1}{k} \cdot q^k + \frac{k+1}{k} \cdot (1 - q^k) = 1 + \frac{1}{k} - q^k$, N 份血样化验次数的平均值即为 $N(1 + \frac{1}{k} - q^k)$.

分组化验模型的核心问题是:

*收稿日期: 2012-11-30

接收日期: 2013-09-03

基金项目: 河北省高等学校科学技术研究项目“血液分组化验模型的优化研究”(Z2013064).

作者简介: 高焕江 (1963-), 男, 河北沧州, 副教授, 研究方向: 医药数理分析与建模.

① 当血样阳性概率 p 为多少时, 分组化验比不分组化验能够减少化验次数^[1,6]? 即 p 值在什么范围时, 存在正整数 $k > 1$ 使得 $f(k) = 1 + \frac{1}{k} - q^k < 1$?

② 当血样阳性概率 p 给定, 如何确定每组的血样份数使化验次数最少^[1,6]? 即 p 值给定, k 取何值时, $f(k) = 1 + \frac{1}{k} - q^k$ 取得最小值?

3 主要结果及证明

3.1 分组化验对血样阳性率的要求

设血样阳性概率为 $p(0 < p < 1)$, 若存在正整数 $k > 1$ 使得 $1 + \frac{1}{k} - q^k < 1$, 当且仅当 $q^k > \frac{1}{k}$, 即 $q > \frac{1}{\sqrt[k]{k}}$. 为此, 首先讨论函数 $g(x) = \frac{1}{\sqrt[x]{x}} (x > 1)$ 的性态.

引理 1^[1] 函数 $g(x) = \frac{1}{\sqrt[x]{x}} (x > 1)$ 当 $x = e$ 时取得最小值 $g(e) = \frac{1}{\sqrt[e]{e}}$.

证 $g'(x) = \frac{1}{\sqrt[x]{x}} \cdot \frac{1}{x^2} (\ln x - 1)$.

令 $g'(x) = 0$, 得 $g(x)$ 的唯一驻点 $x = e$. 当 $1 < x < e$ 时, $g'(x) < 0$; 当 $x > e$ 时, $g'(x) > 0$.

故 $g(x)$ 在 $x = e$ 处取得极小值 (也是最小值) $g(e) = \frac{1}{\sqrt[e]{e}}$.

定理 1 设 $f(k) = 1 + \frac{1}{k} - q^k (0 < q < 1, k \in N_+, k > 1)$, 则当且仅当 $\frac{1}{\sqrt[3]{3}} < q < 1$ 时, 存在大于 1 的正整数 k 使得 $f(k) < 1$.

证 充分性 当 $\frac{1}{\sqrt[3]{3}} < q < 1$ 时, 存在 $k = 3$ 使得 $f(k) = 1 + \frac{1}{k} - q^k = 1 + \frac{1}{3} - q^3 < 1$.

必要性 对给定的血样阳性概率 $p(0 < p < 1)$, 若存在正整数 $k > 1$ 使得 $1 + \frac{1}{k} - q^k < 1$, 则 $q^k > \frac{1}{k}$, 即 $q > \frac{1}{\sqrt[k]{k}}$.

由引理 1 的证明过程可知, $g(x) = \frac{1}{\sqrt[x]{x}}$ 在区间 $(e, +\infty)$ 内单调增大, 故

$$g(3) < g(4) < g(5) < g(6) < \cdots;$$

又因 $g(2) = g(4) = \frac{\sqrt{2}}{2}$, $\lim_{k \rightarrow +\infty} \frac{1}{\sqrt[k]{k}} = 1$, 于是

$$g(3) < g(2) = g(4) < g(5) < g(6) < \cdots < 1,$$

这说明当 x 取大于 1 的正整数时, 对应的整数点的函数值以 $g(3) = \frac{1}{\sqrt[3]{3}}$ 为最小, 故

$$\frac{1}{\sqrt[3]{3}} \leq \frac{1}{\sqrt[k]{k}} < q < 1.$$

证毕.

这个定理说明: 当 $0 < q \leq \frac{1}{\sqrt[3]{3}}$ 时, 无论怎样选取正整数 k 都不能减少化验次数^[1]; 而当 $\frac{1}{\sqrt[3]{3}} < q < 1$ 时一定存在大于 1 的正整数 k 使化验次数减少. 即是说, 当且仅当血液的阳性概率 p 在区间 $(0, 1 - \frac{1}{\sqrt[3]{3}})$ 内时, 通过恰当分组实施分组化验能使化验次数减少 (相对于不分组化验).

需要进一步考虑的问题是: 对于给定的血样阳性概率 $p(0 < p < 1 - \frac{1}{\sqrt[3]{3}})$, 是否存在大于 1 的正整数 k^* 使化验次数最少? 若存在, 如何确定?

以下称使得化验次数最少的分组方案为最佳分组方案, 最佳分组对应的 k 值称作最佳分组 k 值^[1].

3.2 最佳分组 k 值的存在性

引理 2^[1] 函数 $f(x) = 1 + \frac{1}{x} - q^x (x > 1, \frac{1}{\sqrt[3]{3}} < q < 1)$ 存在最小值, 最小值点 $x^* \in (1, \frac{-1}{\ln q})$ (证明见文献 [1]).

这个引理证明了函数 $f(x) = 1 + \frac{1}{x} - q^x (x > 1, \frac{1}{\sqrt[3]{3}} < q < 1)$ 的最小值的存在性. 针对给定的血样阳性概率 $p (0 < p < 1 - \frac{1}{\sqrt[3]{3}})$, 若能求得函数 $f(x) = 1 + \frac{1}{x} - q^x (x > 1, \frac{1}{\sqrt[3]{3}} < q < 1)$ 的最小值点 x^* , 则只需比较 $f(x)$ 在 $[x^*]$ 与 $[x^* + 1]$ 处的函数值即可确定最佳分组 k 值 k^* (二者对应的函数值相等时, 为使分组组数尽可能少最佳分组 k 值取 $[x^* + 1]$).

3.3 最佳分组 k 值的确定方法

下文讨论最佳分组 k 值的确定方法时将用到如下迭代收敛定理:

引理 3^[7] 设在区间 (a, b) 内 $h'(x) > 0$, $h''(x)$ 保持符号, 且 $a < h(a)$, $b > h(b)$, 则对任意初值 $x_0 \in (a, b)$, 由迭代公式 $x_{n+1} = h(x_n) (n = 0, 1, 2, 3, \dots)$ 产生的数列 $\{x_n\}$ 单调收敛于方程 $x = h(x)$ 在 (a, b) 内的唯一实根 x^* (证明见文献 [7]).

以下利用引理 2 和引理 3 证明如下结论:

定理 2 对于任意初值 $x_0 \in (1, \frac{-1}{\ln q})$, 由迭代式 $x_{n+1} = \sqrt{\frac{-1}{q^{x_n} \ln q}}$ 产生的数列 $\{x_n\}$ 必收敛于函数 $f(x) = 1 + \frac{1}{x} - q^x (x > 1, \frac{1}{\sqrt[3]{3}} < q < 1)$ 在区间 $(1, \frac{-1}{\ln q})$ 内的最小值点 x^* .

证 由引理 2 知, 函数 $f(x) = 1 + \frac{1}{x} - q^x (x > 1, \frac{1}{\sqrt[3]{3}} < q < 1)$ 的最小值点 $x^* \in (1, \frac{-1}{\ln q})$, 且 x^* 恰是 $\varphi(x) = 1 + x^2 q^x \ln q$ 在区间 $(1, \frac{-1}{\ln q})$ 内的零点 (因 $f'(x) = -\frac{1}{x^2}(1 + x^2 q^x \ln q)$).

由 $x^2 q^x \ln q + 1 = 0$, 得 $x = \sqrt{\frac{-1}{q^x \ln q}}$.

令迭代函数 $h(x) = \sqrt{\frac{-1}{q^x \ln q}}$, 根据引理 3, 只需证明:

- ① $h(1) > 1, h(\frac{-1}{\ln q}) < \frac{-1}{\ln q}$;
- ② 在区间 $(1, \frac{-1}{\ln q})$ 内 $h'(x) > 0, h''(x)$ 保持符号.

由

$$\frac{1}{\sqrt[3]{e}} < \frac{1}{\sqrt[3]{3}} < q < 1 \Rightarrow -\frac{1}{e} < \ln q < 0 \Rightarrow 0 < -\ln q < \frac{1}{e} \Rightarrow 0 < -q \ln q < \frac{q}{e} < \frac{1}{e} \Rightarrow \frac{1}{-q \ln q} > e,$$

有 $h(1) = \sqrt{\frac{-1}{q \ln q}} > \sqrt{e} > 1$; 由 $0 < -\ln q < \frac{1}{e} \Rightarrow \frac{1}{-\ln q} > e$, 有

$$h(\frac{-1}{\ln q}) = \sqrt{\frac{-1}{q^{\frac{-1}{\ln q}} \cdot \ln q}} = \sqrt{\frac{-1}{\frac{1}{e} \ln q}} = \sqrt{\frac{1}{-\ln q} \cdot e} < \sqrt{\frac{1}{-\ln q} \cdot \frac{1}{-\ln q}} = \frac{1}{-\ln q}.$$

对任意 $x \in (1, \frac{-1}{\ln q})$, 有

$$\begin{aligned} h'(x) &= -\frac{1}{2} \sqrt{\frac{-1}{\ln q}} \cdot q^{-\frac{x}{2}} \cdot \ln q = \frac{1}{2} \sqrt{-\ln q} \cdot q^{-\frac{x}{2}} > 0, \\ h''(x) &= -\frac{1}{4} \sqrt{-\ln q} \cdot q^{-\frac{x}{2}} \ln q = \frac{1}{4} \sqrt{(-\ln q)^3} \cdot q^{-\frac{x}{2}} > 0 \end{aligned}$$

不变号. 证毕.

由 $-\frac{1}{e} < \ln q < 0$, $-\frac{1}{\ln q} > e$, 知 $2 \in (1, \frac{-1}{\ln q})$. 对给定的血样阳性概率 $p (0 < p < 1 - \frac{1}{\sqrt[3]{3}})$, 可取初值 $x_0 = 2$ 进行迭代求得 x^* .

例如, 当 $p = 0.12$ 时, 初始值取 $x_0 = 2$, 迭代公式 $x_{n+1} = \sqrt{\frac{-1}{0.88^{x_n} \ln 0.88}}$, 经迭代求得 $x^* \approx 3.4976$. 于是, 最佳分组 k 值 (设为 k^*) 应为 3 或 4.

由于

$$f(3) = 1 + \frac{1}{3} - 0.88^3 \approx 0.6519, f(4) = 1 + \frac{1}{4} - 0.88^4 \approx 0.6503,$$

故当 $p = 0.12$ 时 $k^* = 4$.

对于给定的血样阳性概率 $p(0 < p < 1 - \frac{1}{\sqrt[3]{3}})$, 表 1 给出了经迭代求出的函数

$$f(x) = 1 + \frac{1}{x} - q^x (x > 1, \frac{1}{\sqrt[3]{3}} < q < 1)$$

的最小值点 x^* , 以及最佳分组 k 值 k^* , 预期每份血样平均化验次数 $f(k^*)$, 平均化验次数降低率 $1 - f(k^*)$. 为下文得出最佳分组 k 值计算公式, 针对每个 p 值表中第二列列出了对应的 $\frac{1}{\sqrt{p}}$ 的值.

表 1 血液分组化验最佳分组情况^[1,6]

p	$\frac{1}{\sqrt{p}}$	x^*	k^*	$f(k^*)$	$1 - f(k^*)$	p	$\frac{1}{\sqrt{p}}$	x^*	k^*	$f(k^*)$	$1 - f(k^*)$
0.0001	100	100.501	101	0.0200	98.00%	0.13	2.77	3.394	3	0.6748	33.52%
0.001	31.62	32.127	32	0.0628	93.72%	0.14	2.67	3.303	3	0.6973	33.27%
0.005	14.14	14.653	15	0.1391	86.09%	0.15	2.58	3.223	3	0.7192	28.08%
0.01	10.00	10.516	11	0.1956	80.44%	0.16	2.5	3.152	3	0.7406	25.94%
0.02	7.07	7.597	8	0.2742	72.58%	0.17	2.43	3.089	3	0.7615	23.85%
0.03	5.77	6.308	6	0.3337	66.63%	0.18	2.36	3.033	3	0.7820	21.80%
0.04	5.00	5.542	6	0.3839	61.61%	0.19	2.29	2.983	3	0.8019	19.81%
0.05	4.47	5.022	5	0.4262	57.38%	0.20	2.24	2.938	3	0.8213	17.87%
0.06	4.08	4.641	5	0.4661	53.39%	0.21	2.18	2.898	3	0.8403	15.97%
0.07	3.78	4.346	4	0.5019	49.81%	0.22	2.13	2.863	3	0.8588	14.12%
0.08	3.54	4.110	4	0.5336	46.64%	0.23	2.09	2.832	3	0.8768	12.32%
0.09	3.33	3.917	4	0.5643	43.57%	0.24	2.04	2.805	3	0.8944	10.56%
0.10	3.16	3.755	4	0.5939	40.61%	1/4	2.00	2.782	3	0.9115	8.85%
0.11	3.02	3.617	4	0.6226	37.74%	0.26	1.96	2.762	3	0.9281	7.19%
1/9	3.00	3.602	4	0.6257	37.43%	0.27	1.92	2.746	3	0.9443	5.57%
0.12	2.89	3.498	4	0.6503	34.97%	0.28	1.89	2.734	3	0.9601	3.99%
0.123	2.85	3.465	4	0.6584	34.16%	0.29	1.86	2.725	3	0.9754	2.46%
0.124	2.84	3.454	3	0.6611	33.89%	0.30	1.83	2.720	3	0.9903	0.97%

3.4 最佳分组 k 值计算公式

表 1 针对给定的先验概率 $p(0 < p < 1 - \frac{1}{\sqrt[3]{3}})$, 用迭代法求出了若干分组化验最佳分组 k 值. 可以看出, 当血样呈阳性的概率不大于 0.1 时, 通过恰当分组实施分组化验可显著减少平均化验次数, 预期平均化验次数降低率达 40% 以上.

观察表 1 中数据并考虑到函数 $\varphi(p) = \frac{1}{\sqrt{p}}$ 在区间 $(0, 1 - \frac{1}{\sqrt[3]{3}})$ 上的单调性可以发现, 对于给定的血样阳性概率 $p(0 < p < 1 - \frac{1}{\sqrt[3]{3}})$, 分组化验最佳分组 k 值呈现一定的规律性. 若能揭示出其中的规律 (而不用针对每个具体的血样阳性概率 p 进行迭代), 则无论是对理论研究还是临床实践都将是十分重要的.

分析表中数据可知, 当血样阳性概率 p 在区间 $(\frac{1}{9}, \frac{1}{4}]$ 上由小到大变化时, 最佳分组 k 值先是取 4 后是取 3, 对应的 p 值分界点应在 0.12 与 0.13 之间. 当 $p = 0.123$ 时最佳分组 k 值是 4, 而当 $p = 0.124$ 时最佳分组 k 值是 3, 二者对应的每份血样平均化验次数相差甚微. 据此推测, 介于 0.123 与 0.124 之间存在一个 p_1^* , 当血样阳性概率为 p_1^* 时, 将总体分为 4 份血样一组和分为 3 份血样一组对应的平均化验次数相等.

事实上, 由

$$f(3) = 1 + \frac{1}{3} - q^3, f(4) = 1 + \frac{1}{4} - q^4,$$

令 $f(3) = f(4)$, 则 $q^3 - q^4 - \frac{1}{12} = 0$, 从而 $q^4 = q^3 - \frac{1}{12}$, 采用迭代格式 $q_{n+1} = \sqrt[4]{q_n^3 - \frac{1}{12}}$, 初值取 $q_0 = 1 - \frac{1}{9} = \frac{8}{9}$, 经迭代求得 $q_1^* \approx 0.8761$, 故分界点 $p_1^* \approx 0.1239$.

同样, 当血样阳性概率 p 在区间 $(\frac{1}{16}, \frac{1}{9}]$ 上由小到大变化时, 最佳分组 k 值先是 5 然后是 4, 由 $f(4) = 1 + \frac{1}{4} - q^4, f(5) = 1 + \frac{1}{5} - q^5$, 令 $f(4) = f(5)$, 得 $q^5 = q^4 - \frac{1}{20}$, 采用迭代格式 $q_{n+1} = \sqrt[5]{q_n^4 - \frac{1}{20}}$, 初值取 $q_0 = 1 - \frac{1}{16} = \frac{15}{16}$, 迭代求得 $q_2^* \approx 0.9344$, 分界点为 $p_2^* \approx 0.0656$.

一般情况是, 当血样阳性概率 p 在区间 $(\frac{1}{(m+1)^2}, \frac{1}{m^2}]$ ($m \geq 2$ 为正整数) 上由小到大变化时, 最佳分组 k 值先是取值 $m+2$, 待 p 值达到某一确定概率点 p^* (分界点) 后, 最佳分组 k 值为 $m+1$, 由迭代格式

$$q_{n+1} = \sqrt[m+2]{q_n^{m+1} - \frac{1}{(m+2)(m+1)}},$$

初值取 $q_0 = 1 - \frac{1}{(m+1)^2}$, 迭代可求 q^* , 分界点 $p^* = 1 - q^*$ 即可确定.

综上, 若令

$$(\frac{1}{4}, 1 - \frac{1}{\sqrt[3]{3}}) \cup \sum_{m=2}^{\infty} \left(\frac{1}{(m+1)^2}, \frac{1}{m^2} \right]$$

为区间 $(0, 1 - \frac{1}{\sqrt[3]{3}})$ 的一个分割, 则分组化验最佳分组 k 值在区间 $(0, 1 - \frac{1}{\sqrt[3]{3}})$ 上是血样阳性概率 p 的阶梯函数, 在任意子区间 $(\frac{1}{(m+1)^2}, \frac{1}{m^2}]$ ($m \geq 2$ 为正整数) 和区间 $(\frac{1}{4}, 1 - \frac{1}{\sqrt[3]{3}})$ 上函数表达式为

$$k^* = \begin{cases} m+2 \text{ 或 } m+1, & \frac{1}{(m+1)^2} < p \leq \frac{1}{m^2} \\ 3, & \frac{1}{4} < p < 1 - \frac{1}{\sqrt[3]{3}} \end{cases} = \begin{cases} m+2, & \frac{1}{(m+1)^2} < p \leq p^*, \\ m+1, & p^* < p \leq \frac{1}{m^2}, \\ 3, & \frac{1}{4} < p < 1 - \frac{1}{\sqrt[3]{3}}, \end{cases}$$

其中 $m \geq 2$ 为正整数, $p^* = 1 - q^*$, q^* 可由迭代格式

$$q_{n+1} = \sqrt[m+2]{q_n^{m+1} - \frac{1}{(m+2)(m+1)}},$$

初值取 $q_0 = 1 - \frac{1}{(m+1)^2}$ 迭代求得 (迭代足够次数可使结果达到要求的精度).

4 结语

本文研究结论表明, 当血样呈阳性的概率不大于 0.1 时, 通过恰当分组实施分组化验可显著降低平均化验次数, 预期平均化验次数降低率达 40% 以上, 从而大幅度节约时间, 减少人力 (物力) 成本提高工作效率和社会效益. 本文从数理上证明了当血样呈阳性的概率 p 在某

一确定范围时分组化验存在最佳分组方案, 针对给定的血样阳性概率 p 给出了最佳分组 k 值的确定方法, 得出了分组化验最佳分组 k 值计算公式.

参 考 文 献

- [1] 高焕江. 关于分组化验模型的一个猜想 [J]. 数学的实践与认识, 2012, 42(20) : 135–138.
- [2] 王为. 分组检验成立的条件及最佳分组方法 [J]. 新疆工学院学报, 1996, 17(1) : 19–23.
- [3] 张贵生. 分组验血法最佳分组人数的取值范围和估计公式 [J]. 苏州医学院学报, 1994, 14(6) : 567–570.
- [4] 张乐成, 马跃, 赵旭. 血液分组化验问题二次分组化验法最佳分组方式的算法 [J]. 数学的实践与认识, 2011, 41(1) : 78–84.
- [5] 毕义明, 张红文, 杨宝珍. 血样二次分组检验的探讨 [J]. 数学的实践与认识, 2010, 40(24) : 88–91.
- [6] 高焕江. 血液化验二次分组模型的简捷算法 [J]. 数学的实践与认识, 2013, 43(4) : 53–59.
- [7] 周国才, 徐荣良, 孙昭. 一点迭代法的收敛定理 [J]. 太原工业大学学报, 1986, (1) : 9–12.

STUDY ON OPTIMIZATION OF GROUP TEST MODEL

GAO Huan-jiang ¹, ZHANG Cui-li ¹, GAO Peng ²

(1. Department of Occupational Mathematics, Xingtai Medical College, Xingtai 054000, China)

(2. College of Preventive Medicine, Third Military Medical University, Chongqing 400038, China)

Abstract: In this paper, we consider a model of blood group test. The existence for the optimum method of grouping is proved when the positive probability of blood sample lies in a definite region by using probability theory and differential calculus. It puts forward a method and derives a formula in questing of the optimum number k of group sample members with the definite positive probability.

Keywords: group test; mathematical expectation; optimum number k of group sample members; iterative formula; step function

2010 MR Subject Classification: 60H30