

人类细胞遗传学研究的数学方法

杜荣骞

(南开大学生物系,天津)

做人类染色体核型分析时,最感困难的是如何正确区分嵌合体与正常二倍体。在统计学中,如果将正常人误诊为病人,称为犯 I 型错误;如果将病人误诊为正常人,称为犯 II 型错误。为了能够用最低的工作量,得到可靠的实验结果,并将实验误差控制在一定的范围内,本文提出序贯分析方法,供科研及临床实践参考。

原 理

序贯检验方法^[1]是 Wald 以概率比检验为基础提出来的^[2]。用该方法检验时,需要知道正常人非整倍体细胞的比率 p_0 及嵌合体非整倍体细胞的比率 p_1 。对于一个未知的个体,当其非整倍体细胞出现的比率 $p = p_0$ 时,则以概率 $1-\alpha$ 认为是正常人 (α 为犯 I 型错误的概率);当 $p = p_1$ 时,则以概率 β 认为是正常人 (β 为犯 II 型错误的概率)。当把 α 和 β 定得适当小以后,我们就会以很低的概率将正常人误诊为病人或将病人误诊为正常人。

每做一次观察之后,都会得出一个 p , 然后做 $p = p_0$ 时的概率 $P(0)$ 与 $p = p_1$ 时的概率 $P(1)$ 之间的比较。若 $R = P(1)/P(0)$ 很大时,则认为 $p = p_1$ 的可能性远远大于 $p = p_0$ 的可能性。这时拒绝 $p = p_0$, 认为该个体为嵌合体。若 R 很小,则认为 $p = p_1$ 的可能性远远小于 $p = p_0$ 的可能性。这时接受 $p = p_0$, 认为该个体为正常人。若 R 不太大也不太小, 这时既不拒绝 $p = p_0$, 也不接受 $p = p_0$, 再做下一次观察。

可以用 A 、 B 定出两条界:

当 $R \leq B$ 时,接受 $p = p_0$, 由 $R \leq B$ 所确定的区域称为接受区域。

当 $R \geq A$ 时,拒绝 $p = p_0$, 由 $R \geq A$ 所确定的区域称为拒绝区域。

当 $B < R < A$ 时,不能做出接受或拒绝的决定。由 $B < R < A$ 所决定的区域称为不定区域。

当 $B < R_1 < A$ 时,需做第二次观察。在观察下一个细胞后,得到 R_2 。若 $R_2 \leq B$, 则接受 $p = p_0$; 若 $R_2 \geq A$, 则拒绝 $p = p_0$; 若 $B < R_2 < A$, 则不做决定, 再做第三次观察。如此进行下去,直到 $R_i \leq B$ 或 $R_i \geq A$, 做出相应的接受或拒绝 $p = p_0$ 时为止。其中的 A 、 B 可由以下二式决定。

$$A = \frac{1-\beta}{\alpha}, \quad B = \frac{\beta}{1-\alpha}。$$

设做了 i 次抽样, 共观察了 n 个细胞, 其中含有 k 个异常细胞, 且不能做出接受或拒绝 $p = p_0$ 的决定时, k 应满足:

$$B < \frac{p_1^k (1-p_1)^{n-k}}{p_0^k (1-p_0)^{n-k}} < A,$$

取对数,

$$\lg B < k \lg \frac{p_1}{p_0} + (n-k) \lg \frac{1-p_1}{1-p_0} < \lg A,$$

整理为:

$$\frac{\lg B}{\lg \frac{p_1}{p_0} - \lg \frac{1-p_1}{1-p_0}} < k < \frac{\lg A}{\lg \frac{p_1}{p_0} - \lg \frac{1-p_1}{1-p_0}}$$

$$- \frac{n \lg \frac{1-p_1}{1-p_0}}{\lg \frac{p_1}{p_0} - \lg \frac{1-p_1}{1-p_0}},$$

令:

$$b = \frac{\lg B}{\lg \frac{p_1}{p_0} - \lg \frac{1-p_1}{1-p_0}},$$

$$a = \frac{\lg A}{\lg \frac{p_1}{p_0} - \lg \frac{1-p_1}{1-p_0}},$$

$$m = \frac{-\lg \frac{1-p_1}{1-p_0}}{\lg \frac{p_1}{p_0} - \lg \frac{1-p_1}{1-p_0}},$$

则 $b + mn < k < a + mn$ 。
其中 n 为第 i 次观察时的样本含量,以

$$a_n = b + mn \text{ 表示接受阈值,}$$

$$r_n = a + mn \text{ 表示拒绝阈值,}$$

于是得到两个界。在做 i 次观察后,若观察的结果小于 a_n ,落在接受域中,则接受 $p = p_0$;若大于 r_n ,落在拒绝域中,则接受 $p = p_1$;若在 a_n 及 r_n 之间,落在不定域中,则继续观察。

核型分析

正常人类细胞在有丝分裂过程中,由于随机误差,可能会造成一定比例的非整倍体细胞。非整倍体细胞在不同年龄组中是不同的。一般来说,儿童及老年人的比例高,中年及青年人的比例低^[2,3,5]。现将 Court Brown 等人及 Hamerton 等人的资料略加调整列在表 1 中。

由于不同性别,不同年龄个体的亚二倍体及超二倍体细胞的百分率不同。因此,在做序贯分析时应分别不同的情况定出相应的 a_n 及 r_n 。表 1 中的百分率是 p_0 。根据我们在实际应用中的体会,令 $p_1 = 2p_0$ 较为合理。 α 和 β 可以根据实验的要求定为任意小的值,例如

$$\alpha = \beta = 0.05。$$

下面是我们对 0—14 岁年龄组女性受检者的分析。从表 1 中查出 $p_0 = 0.059$, 则

$$p_1 = 2p_0 = 0.118。$$

定 $\alpha = \beta = 0.05$, 计算出:

$$A = \frac{1 - 0.05}{0.05} = 19.0000,$$

$$B = \frac{0.05}{1 - 0.05} = 0.0526。$$

表 1 不同年龄组非整倍体细胞率

性别	年龄组(岁)	亚二倍体(%)	超二倍体(%)
女	0—14	5.9	0.7
	15—54	4.9	0.5
	55+	11.1	1.3
男	0—14	5.1	1.4
	15—54	5.1	0.6
	55+	6.2	1.6

及 $b = -3.885$, $a = 3.885$, $m = 0.085$ 。

因此: $a_n = -3.885 + 0.085n$,

$$r_n = 3.885 + 0.085n。$$

我们分析时采用作图法。横轴为镜检的总细胞数,纵轴为非整倍体细胞数。在座标平面内作出 a_n 及 r_n 两条直线。这两条直线将平面分成三个部分:拒绝域、不定域及接受域(图 1)

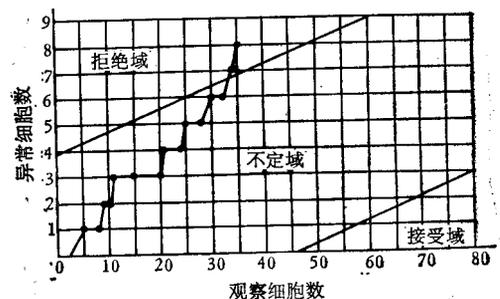


图 1 拒绝域、不定域与接受域的坐标

图 1 是我们分析过的一个病例。每观察一次在图上点上一个点,如观察到 11 个细胞时,出现 3 个异常细胞,则有点 [11,3] 等。将所有点联结起来,则成为连续的折线。当折线在不确定域内游动时,需继续观察。一旦游出不定域而进入接受域,则接受 $p = p_0$;若进入拒绝域,则接受 $p = p_1$,这时即可终止观察。本例在观察到 35 个细胞时,有 8 个亚二倍体细胞,折线进入拒绝域,接受 $p = p_1$ 。结论是:受检者为病人。错误推断的风险,即犯 I 型错误的风险为 0.05。

(参考文献见第 17 页)

化的情况,即在细胞里不见细胞核,整个细胞中充满微核。这种情况可能是由于温度的急剧变化引起染色体的较多断裂而又不能愈合,可能同时也阻碍了纺锤体的形成而使染色体分散在细胞中,形成微核。

在不形成微核的四分体细胞中有时还可以观察到核内形成透明的不着色的空泡化小区的现象,这种现象在激光引起水稻染色体产生畸变的实验中也有所发现^[1],说明温度急剧变化的作用和激光产生的热效应有其一致性。但出现这种现象的原因尚不清楚。

本实验中所观察到的冬季温室中紫露草出现高微核率的情况值得引起注意。根据多年来的实验观察,除了射线外,未见到能引起紫露草四分体微核率在30%以上的诱变因子。一些

强化学诱变剂在高浓度下,常由于致死花粉母细胞的作用,一般也看不到出现很高微核率的情况。

温度急剧变化与突变关系的研究在实验生态学和细胞遗传学方面都有一定的理论意义。本文仅是初步观察的结果。关于温度急剧变化诱发突变的机理等问题有待进一步研究。

参 考 文 献

- [1] 方宗熙: 1979. 普通遗传学, 科学出版社, 175.
 [2] 罗士韦等: 1983. 细胞生物学杂志, (1): 1—7.
 [3] 侯家龙: 1982. 生态学杂志, (2): 43—44.
 [4] 侯家龙等: 1984. 山东海洋学院学报, (2): 56—59.
 [5] 中山大学生物系植生遗传教研室: 1976. 激光, (2): 26—31.
 [6] Ma T. H.: 1981. *Environmental Health Perspectives*, Vol.37, 85—90.

(上接第3页)

参 考 文 献

- [1] Adelberg, E. A. et al.: 1965. *Biochem. Biophys. Res. Commun.*
 [2] Andrew, W. B. et al.: 1978. *J. Gen. Microbiol.*, 104: 209—218.
 [3] Beringer, J. E.: 1974. *ibid.*, 84: 188—189.
 [4] Beringer, J. E. et al.: 1976. *Nature*, 264—294.
 [5] Berman, M. L. et al.: 1982. *Advanced Bacterial Genetics*, New York, Cold Spring Harbor Laboratory.
 [6] Casadesus, J. et al.: 1979. *Mol. Gen. Genet.*, 174:203—209.
 [7] Dean, H. F. et al.: 1983. *J. Bacteriol.*, 153:

485—497.

- [8] Gijsegem, F. V. et al.: 1982. *Plasmid*, 7: 30—34.
 [9] Holloway, B. W.: 1965. *Virology*, 25: 634—642.
 [10] ————: 1975. *Genetics and Biochemistry of Pseudomonas* (P. H. Clarke and M. H. Richmond, eds.), 133—161, Wiley, London.
 [11] Holloway, B. W.: 1978. *J. Bacteriol.*, 133: 1078—1082.
 [12] Lennox, E. S.: 1955. *Virology*, 1: 190—206.
 [13] Megias, M. et al.: 1982. *J. Bacteriol.*, 149: 59—64.
 [14] Pemberton, J. M. et al.: 1981. *ibid.*, 147: 110—117.
 [15] Sistro, W. R. et al.: 1977. *ibid.*, 131: 536—532.

(上接第35页)

参 考 文 献

- [1] 中国科学院数学研究所统计组: 1978. 抽样检验方法, 科学出版社, 95—103.
 [2] Court Brown, W. M. et al.: 1966. *Chromosome*

Study on Adults, Eugenics Laboratory Memoris, London, Cambridge University Press, 42.

- [3] Hamerton, J. L. et al.: 1965. *Nature Lond.*, 206: 1232.
 [4] Wold, A.: 1947. *Sequential Analysis*, New York, John Wiley and Son, Inc.
 [5] Кузнецов, А. И.: 1980 *Генетика*, 16(7): 1285.