

遗传流行病学双生子法中简易遗传度的计算及其统计推断

辽宁省卫生防疫站

章扬熙

在遗传流行病的研究中,双生子法是常用的重要方法。双生子有两类,即单卵双生(monozygotic twin, MZ)和异卵双生(dizygotic twin, DZ)。单卵双生系一个受精卵发育成的双生子,它们的遗传基础相同;发育环境则可能存在差别,异卵(双卵)双生则为两个受精卵发育成的双生子,他们的遗传基础不同(差异程度与全同胞一致,有1/2的基因型相同),发育环境也有差异。遗传是内因,环境是外因,这两个因子是引起生物差异、疾病发生的两个基本因素。在多数情况下,二者各起着部分的作用。为了度量遗传与环境对某一生物差异或疾病发生所起的作用大小,需要一个尺度来度量它,这个尺度即遗传度(heritability),记为 h^2 ,它的定义^[1]是

$$\text{遗传度}(h^2) = \frac{\text{因遗传所引起的差异}}{\text{(遗传+环境)所引起的差异}}$$

单卵双生的差异源于环境,异卵双生的差异源于遗传与环境,基于这个特点,应用双生子来研究生物差异、疾病发生的遗传度则为重要手段。

有两类的遗传度,一类是计数资料的遗传度,一类是计量资料的遗传度。对双生子法遗传度的计算虽有多种,但对其参数估计与假设检验则均未见报道。本文提出一种简易的遗传度计算方法,并给出了参数估计与假设检验的统计推断法。

计数资料的遗传度

Holzinger^[2]曾提出了一个双生子法计数资料遗传度的计算方式为

$$h^2 = \frac{C_{MZ} - C_{DZ}}{1 - C_{DZ}} \quad (\text{公式1})$$

式中 C_{MZ} 和 C_{DZ} 分别为MZ和DZ组的发病一致率(Concordance)。所谓发病一致率是指双生子中一个患某种病,另一个也患同样疾病的发生频率。这个公式的原理是显见的。把双生子发生某病的可能结果,概括为两个互不相容的事件,即发病一致率和非发病一致率。把双卵双生子的非发病一致率 \bar{C}_{DZ} 作为(遗传+环境)所引起的差异,把 $\bar{C}_{DZ} - \bar{C}_{MZ}$ 作为因遗传所引起的差异,则有

$$h^2 = \frac{\bar{C}_{DZ} - \bar{C}_{MZ}}{\bar{C}_{DZ}}$$

由于 $C_{DZ} + \bar{C}_{DZ} = 1$,则 $\bar{C}_{DZ} = 1 - C_{DZ}$;同理, $\bar{C}_{MZ} = 1 - C_{MZ}$,将有关值代入上式

$$\begin{aligned} h^2 &= \frac{(1 - C_{DZ}) - (1 - C_{MZ})}{1 - C_{DZ}} \\ &= \frac{C_{MZ} - C_{DZ}}{1 - C_{DZ}} \end{aligned}$$

例1 在精神分裂症的调查中,30对单卵双生子中共同发生该病的有24对,26对双卵双生子中共同发病的有3对。求精神分裂症的遗传度。

$$C_{MZ} = \frac{24}{30} = 0.80$$

$$C_{DZ} = \frac{3}{26} = 0.12$$

将所得值代入公式1,得

$$h = \frac{0.80 - 0.12}{1 - 0.12} = 0.77$$

说明精神分裂症的遗传度为77%。

计数资料遗传度的统计推断

1. 对遗传度的假设检验: 上例求得的遗传度0.77是样本的遗传度, 这个遗传度会不会是从总体遗传度为零中抽取的呢? 可以进行 χ^2 检验, 以例1为例, 求 χ^2 值

$$\chi^2 = \frac{(|ad - bc| - 0.5)^2 N}{(a+b)(c+d)(a+c)(b+d)}$$

$$= \frac{(|24 \times 23 - 6 \times 3| - 0.5)^2 \times 56}{30 \times 26 \times 27 \times 29}$$

$$= 26.10$$

$$S_h^2 = \sqrt{\left[\frac{1 - C_{MZ}}{(1 - C_{DZ})^2} \right]^2 \cdot \frac{C_{DZ}(1 - C_{DZ})}{n_2} + \left[\frac{1}{1 - C_{DZ}} \right]^2 \cdot \frac{C_{MZ}(1 - C_{MZ})}{n_1}} \quad (\text{公式4})$$

式中 n_1 为单卵双生子对数, n_2 为双卵双生子对数。

例2 调查303对双生子, 单卵双生201对中近视一文者156对, 双卵双生102对中近视一致者47对, 求遗传度及其95%可信限。

$$C_{MZ} = \frac{156}{201} = 0.78$$

$$S_h^2 = \sqrt{\left[\frac{1 - 0.78}{(1 - 0.46)^2} \right]^2 \cdot \frac{0.46(1 - 0.46)}{102} + \left(\frac{1}{1 - 0.46} \right)^2 \cdot \frac{0.78(1 - 0.78)}{201}} = 0.0657$$

应用公式2求总体 h^2 的95%可信限为
 $0.59 \pm 1.96 \times 0.0657 = 0.46 \sim 0.72$

3. 假设检验: 有的场合需要对两个样本遗传度进行比较。在较大样本时, 可进行u检验。公式为

$$u = \frac{|h_1^2 - h_2^2|}{\sqrt{S_{h_1}^2 + S_{h_2}^2}} \quad (\text{公式5})$$

式中 h_1^2 、 h_2^2 为两个比较的遗传度, $S_{h_1}^2$ 、 $S_{h_2}^2$ 分别为这两个遗传度标准误的平方, 其标准误用公式4计算。

计量资料的遗传度

双生子法计量资料遗传度的计算方法很多, 经权衡选一简易且较好的方法。Von verschuer^[3]提出了用双卵双生的平均相对偏差与单卵双生的平均相对偏差的比值来考察遗传作用的大小。先应用下式计算每对双生的相对偏差D

自由度 $df=1$, 界值 $\chi_{0.05}^2=3.84$, $\chi_{0.01}^2=6.63$, 故 $P < 0.01$ 说明两组发病一致率有非常显著的差别, 即遗传度0.77不是从总体遗传度为零的总体中抽取的。

2. 参数估计: 在较大的样本中总体遗传度的95%和99%可信限依次从下式求得

$$h^2 \pm 1.96 S_h^2 \quad (\text{公式2})$$

$$h^2 \pm 2.58 S_h^2 \quad (\text{公式3})$$

式中 S_h^2 为遗传度的标准误, 依方差传播定律推导得下式

$$C_{DZ} = \frac{47}{102} = 0.46$$

$$h^2 = \frac{0.78 - 0.46}{1 - 0.46} = 0.59$$

原题给出 $n_1=201$, $n_2=102$, 将有关值代入公式4, 得

$$D = \frac{|a - b|}{a + b} \times 100 \quad (\text{公式6})$$

再对每组求平均相对偏差 \bar{D} , 公式为

$$\bar{D} = \frac{\sum D}{n} \quad (\text{公式7})$$

公式6中a和b为一对双生子的两个计量值, 公式7中n为双生子对数。

$\bar{D}_{DZ} : \bar{D}_{MZ}$ 即其比值。比值等于1说明无遗传作用, 比值越大说明遗传作用越大。这个比值可视为遗传指数 I_h , 它与遗传度的概念不同。根据遗传度的定义, 可用下式计算遗传度。

$$h^2 = \frac{\bar{D}_{DZ} - \bar{D}_{MZ}}{\bar{D}_{DZ}} \quad (\text{公式8})$$

例3 某卫生防疫站测量73对双卵双生的身长和208对单卵双生的身长得一计量资料, 求其遗传度。

先应用公式6求每对双生的相对偏差D, 比如单卵双生的第一对值为87厘米和84厘米,

$$D = \frac{|87 - 84|}{87 + 84} \times 100 = 1.75$$

余类推。将两组的D值分别制作频数表如下

附表 MZ组与DZ组D值的频数

组段	组中值x	MZ组的频数 f_1	DZ组的频数 f_2
0—	0.25	108	16
0.5—	0.75	58	7
1.0—	1.25	27	6
1.5—	1.75	12	8
2.0—	2.25	3	9
2.5—	2.75		8
3.0—	3.25		9
3.5—	3.75		4
4.0—	4.25		2
4.5—	4.75		1
5.0—	5.25		3
合计		208	73

$$\bar{D}_{MZ} = \frac{\sum f_1 x}{n_1} = \frac{132}{208} = 0.6346$$

$$\bar{D}_{DZ} = \frac{\sum f_2 x}{n_2} = \frac{146.25}{73} = 2.0034$$

再用公式 8 求遗传度，得

$$h^2 = \frac{2.0034 - 0.6346}{2.0034} = 0.68$$

计量资料遗传度的统计推断

1. 对遗传度的假设检验：一个样本的遗传度有无意义，即会不会从总体遗传度为零中抽取的，可对其作假设检验。由于D的分布呈偏态，小样本时可用秩和检验，对DZ与MZ两组

$$S_h^2 = \sqrt{\left(\frac{\bar{D}_{MZ}}{\bar{D}_{DZ}^2}\right)^2 \cdot \left(\frac{S_{DZ}^2}{n^2}\right) + \left(\frac{1}{\bar{D}_{DZ}}\right)^2 \cdot \left(\frac{S_{MZ}^2}{n_1}\right)} \quad (\text{公式 9})$$

例 4 求例 3 总体遗传度 95% 的可信限

先用公式 9 求 S_h^2

$$S_h^2 = \sqrt{\left[\frac{0.6346}{(2.0034)^2}\right]^2 \cdot \frac{(1.4170)^2}{73} + \left(\frac{1}{2.0034}\right)^2 \cdot \frac{(0.4901)^2}{208}} = 0.03123$$

将有关结果代入公式 2，得总体 h^2 的 95% 可信限为

$$0.68 \pm 1.96 \times 0.03123 = 0.62 \sim 0.74$$

3. 假设检验：两个大样本的遗传度的比较，可用 u 检验，公式仍用公式 5，但式中 S_h^2 则用公式 9 求得 S_h 后再平方。

资料进行比较。根据中心极限定理，在大样本时 \bar{D} 近似正态分布^[4]，所以可对其作 u 检验。

以例 3 为例来具体运算。

先求标准差

$$S_{MZ} = \sqrt{\frac{\sum f_1 x^2 - \frac{(\sum f_1 x)^2}{n_1}}{n_1 - 1}}$$

$$= \sqrt{\frac{133.5 - \frac{(132)^2}{208}}{208 - 1}} = 0.4901$$

$$S_{DZ} = \sqrt{\frac{\sum f_2 x^2 - \frac{(\sum f_2 x)^2}{n_2}}{n_2 - 1}}$$

$$= \sqrt{\frac{437.56 - \frac{(146.25)^2}{73}}{73 - 1}} = 1.4170$$

再求 u 值

$$u = \frac{|\bar{D}_{DZ} - \bar{D}_{MZ}|}{\sqrt{\frac{S_{MZ}^2}{n_1} + \frac{S_{DZ}^2}{n_2}}}$$

$$= \frac{2.0034 - 0.6346}{\sqrt{\frac{(0.4901)^2}{208} + \frac{(1.4170)^2}{73}}} = 8.09$$

界值 $u_{0.05} = 1.96$ ， $u_{0.01} = 2.58$ ，所以 $P < 0.01$ ，说明两组平均相对偏差的差别显著， h^2 有意义。

2. 参数估计：对大样本总体遗传度的 95% 及 99% 可信限仍用公式 2 与公式 3 计算。惟式中的遗传度标准误 S_{h^2} ，依方差传播定律推导得下式

例 5 某卫生防疫站对 281 对男性双生、213 对女性双生测量身长，男性 $h^2_1 = 0.68$ ， $S_{h^2_1} = 0.03123$ ，女性 $h^2_2 = 0.72$ ， $S_{h^2_2} = 0.03547$ ，试比较两个遗传度有无显著差别？

将有关数值代入公式 5 求 u 值

$$u = \frac{|0.72 - 0.68|}{\sqrt{0.03123^2 + 0.03547^2}} = 0.85$$

u的界值 $u_{0.05} = 1.96$, 所以 $P > 0.05$, 说明男女身长的遗传度未见显著差别。

讨 论

1. 关于双生子法遗传度的计算公式, 在计数资料方面较为统一; 但在计量资料方面则有多种, 概括起来, 有三类。一类是用相对偏差反映差异, 一类是用 $1-r$ (相关系数) 来反映差异, 一类是用方差来反映差异。应当指出, 相关不等于相同, $r=1$ 不等于无差异, 用 $1-r$ 反映差异往往求得 h^2 偏高, 值得商榷。应用方差分析有几个基本假定, 即(1)处理效应与环境效应可加, (2)实验各组数据的总体呈正态分布, 且有相同的方差。这些条件在双生子法的数据中, 未必能完全满足, 且方差分析的参数估计并非像有些用其计算 h^2 公式那样简单, 不但计算较繁, 且统计推断也较难。用相对偏差来反映差异较为简易, 只要样本较大, 根据中心极限定理等理论可知 \bar{D} 分布与 h^2 分布近似正态, 这就为统计推断带来了方便。

2. 双生子法计量资料遗传度不同公式的计算结果是否一致? 用组内相关系数求得的 h^2 偏高, 其缘由已述及。平均相对偏差与变异系数近似, $\Sigma(|a-b|)/n$ 约为一个标准差, 它的结果

通常比一个方差要小, 所以往往用平均相对偏差求得的 h^2 值比用方差计算的 h^2 值略小些, 这说明几种公式求得 h^2 值不一致是有缘由的, 对所得 h^2 的意义和可靠性、正确性的估计应依赖于假设检验和参数估计的结果。

摘 要

本文研究了一种新的双生子法简易遗传度的计算及其统计推断的方法, 并对计量资料遗传度不同公式计算结果的不一致性原因进行了讨论。

ABSTRACT

This paper studied a new simple method for calculating the heritability and its statistical inference in twin births study of genetic epidemiology. Besides, the causes of disagreement of some heritabilities by using different equations in one measurement data were discussed.

参 考 文 献

1. 杨纪珂等: 应用生物统计, 第一版, 北京: 科学出版社, 388~397页, 1983
2. 钱宇平等: 流行病学进展, 第三卷, 第一版, 北京: 人民卫生出版社, 250~261页, 1985
3. 铃木尚, 江原昭信: 双生儿の生体计测。文部省总合研究报告, No 12: 50, 1954
4. DeGroot MH: Probability and Statistics California: Addison-wesley Publishing Company Inc, p 157~160, 1975

南昌地区婴幼儿秋冬季腹泻病因的研究

伍学洲* 潘达鑫* 梅魁敏* 陈志军* 文运弟* 陈六英* 陈德怀** 胡桂珍**

为了研究南昌地区秋冬季婴幼儿腹泻与轮状病毒的关系, 我们于1981和82年秋冬, 自江西省儿童医院临床诊断为秋冬季腹泻的84例患儿粪便标本中取78份作了检查, 结果为:

78份粪便标本经低速离心后取上清液直接滴膜负染, 电镜观察及经免疫电镜观察, 发现有轮状病毒颗粒的阳性标本60份, 阳性检出率为76.9%。

取34例患儿双份血清(第一份血清为患儿入院

时, 第二份为出院后1~6个月采取)作补结抗体试验结果, 4倍增长率为55.9%, 其中28例的结果与电镜检查结果比较, 两者符合率为60.7%。

根据流行季节、临床表现和粪便特征, 认为南昌地区1981和82年秋冬季婴幼儿腹泻的主要病原是轮状病毒。

*江西中医学院 *江西儿童医院

**江西省医科所