

应用程序计算器作流行病学分析的方法与实例

同济医科大学流行病学教研室 何尚浦

流行病学原理与方法近年来有很大发展,并广泛应用于临床医学与预防医学各学科的研究中。资料的收集、归纳与分析是流行病学方法的重要内容。归纳与分析的内容有简单的也有复杂的。在归纳与分析的手段上可由简单的只用笔与纸张即可完成,到需要使用计算机进行复杂的运算。

在选用统计分析方法时,流行病学医师常遇到两类统计分析方法:一类方法是理论上应优先选用,但计算较困难;另一类在理论上可取,虽然不很理想,但计算较为简单。流行病学医师多常规地应用后一类的统计分析方法。随着计算机技术的发展与应用,上述情况将发生根本变化。

在我国微型计算机技术的应用已日益广泛,但微型计算机价格还较贵,不便于携带。一些较好的程序计算器(programmable calculator)或袖珍计算机(pocket computer)有其优点:①体积小,便于携带;②价格较便宜;③有些程序计算器可将程序储存于磁卡上,使用更为方便;④已能进行较复杂的计算。因此,目前一些高级程序计算器或袖珍计算机足以促进流行病学资料分析的革命,它能提供快速、准确与理论上较好的分析结果。

在国内市场上能购买到的程序计算器或袖珍计算机有:TI-59,HP41CV,Casio fx-702,Sharp PC-1500,Casio PB-700等。前两种是用磁卡储存,后两种采用BASIC语言编制程序并以磁带储存程序与资料。各有其优缺点。

本文主要介绍以HP41CV程序计算器作流行病学分析的基本内容、所用公式,并用实例说明^[1]。所用的实例只作为计算的例子来介绍,而不是介绍其流行病学分析的结论。

病例对照研究资料的分析

病例对照研究的资料可以根据病例和对照是否有暴露史归纳成 2×2 表进行分析。为了控制混杂因素,资料还需分层归纳成多个 2×2 表。如果暴露史是分级而且分层归纳,则资料便归纳成多个 $2 \times R$ 表来分析。在配比病例对照研究中,有时以1个病例配R个对照。这些资料的处理与分析较为复杂,但应用程序计算器来处理与计算则很方便。

一、病例对照研究分层 2×2 表资料的分析:病例对照研究的资料先分层归纳成多个 2×2 表(表1)。

表1 病例对照研究资料归纳表

	有暴露史	无暴露史	合计
病例	a_i	c_i	M_{1i}
对照	b_i	d_i	M_{0i}
合计	N_{1i}	N_{0i}	T_i

每个 2×2 表可单独分析,方法见一般流行病学书籍中,这里着重介绍多个 2×2 表资料的综合分析方法,并以实例说明。

1. 假设检验:用Mantel和Haenszel介绍的公式计算 $\chi_{M-H}^{[2]}$ 。

$$\chi_{M-H} = \frac{\sum a_i - \sum M_{1i} N_{1i} / T_i}{\left\{ \sum \left\{ \frac{M_{1i} M_{0i} N_{1i} N_{0i}}{T_i^2 (T_i - 1)} \right\} \right\}^{1/2}} \quad (\text{公式-1})$$

式中 i 为第 i 层。其概率(p)可以正态分布表查得或用程序计算器直接求出。

2. 比值比(odds ratio, R)点估计:如各层的比值比较一致,可以计算分层后总的比值比。

①计算 R_{ML} :这是总比值比的渐近极大似

然估计。应用公式-2和公式-3以迭代求得。

$$\sum a_i - \sum E(a_i|R) = 0 \quad (\text{公式-2})$$

$$R = \frac{E(a_i|R) \{M_{0i} - N_{1i} + E(a_i|R)\}}{\{M_{1i} - E(a_i|R)\} \{N_{1i} - E(a_i|R)\}} \quad (\text{公式-3})$$

式中 $E(a_i|R)$ 为以比值比 R 时 a_i 的预期值。

②计算 R_{M-H} : 以Mantel和Haenszel介绍的方法计算 R_{M-H} 较为简便^[2]。 R_{M-H} 一般较接近 R_{ML} , 在计算时不用迭代, 故较常用(见公式-4):

$$R_{M-H} = \frac{\sum \{a_i d_i / T_i\}}{\sum \{b_i c_i / T_i\}} \quad (\text{公式-4})$$

3. 比值比区间估计: 求比值比区间估计可用公式-5和公式-6^[3]。

$$R = R(1-z) X_{M-H} \quad (\text{公式-5})$$

$$\bar{R} = R^2 / R = R(1+z / \chi_{M-H})$$

(公式-6)

式中 R 为比值比点估计。一般用未经修正的 χ 。求90%可信限时, $Z = 1.645$, 求95%可信限时, $Z = 1.96$ 。 \underline{R} 为比值比下限, \bar{R} 为比值比上限。

4. 非均匀性卡方检验: 以非均匀性卡方检验:(Heterogeneity chisquare test) 来分析各层比值比是否一致。如明显不一致, 则不宜以总比值比来作点估计。这时可以用各层的比值比来分析; 如需作总比值比估计, 则需计算SMR作为标化比值比来分析。

非均匀性卡方检验用公式-7求得。这是渐近似然比检验。

$$\chi^2 = -2 \left[\sum \{a_i \ln \left(\frac{\hat{a}_i}{a_i} \right) + b_i \ln \left(\frac{\hat{b}_i}{b_i} \right) + c_i \ln \left(\frac{\hat{c}_i}{c_i} \right) + d_i \ln \left(\frac{\hat{d}_i}{d_i} \right) \right] \quad (\text{公式-7})$$

式中 \hat{a}_i , \hat{b}_i , \hat{c}_i 和 \hat{d}_i 是以每个表固定的边缘总值又满足公式-2和公式-3时在表 i 的 a , b , c 和 d 值。其自由度为 $n-1$ 。

5. SMR的计算: SMR是以有暴露者为标准。以各层 a 值和预期 a 值总计之比来说明。用公式-8来计算。

$$MSR = \sum a_i / \sum \{b_i c / d_i\}, \quad (\text{公式-8})$$

(实例1): Doll和Hill(1954)报道了吸烟史与肺癌关系研究的资料^[4]。其性别分层资料见表2。

表2 不同性别吸烟史与肺癌关系的研究

	男性吸烟史			女性吸烟史		
	有	无	合计	有	无	合计
病例	647	2	649	41	19	60
对照	622	27	649	28	32	60
合计	1269	29	1298	69	51	120
比值比	14.04			2.47		

将上述有关公式编成程序, 可以很简便获得有关分析的数值。

①假设检验: $\chi_{M-H} = 4.992$, $P = 3.0 \times 10^{-7}$ 说明吸烟与肺癌有明显联系。

②比值比点估计: $R_{M-H} = 4.524$, $R_{ML} = 4.255$, $SMR = 10.973$ 。

③比值比区间估计: R_{ML} 95%可信限为2.410~7.513, R_{M-H} 比值比95%可信限为2.501~8.183。

④非均匀性卡方检验: $\chi^2_{HET}(L, R) = 5.737$, 自由度 = $2 - 1 = 1$ 。 $P = 0.0166$ 。说明男女两层的比值比不一致。不宜用 R_{M-H} 、或 R_{ML} 来总的估计其比值比。可分别分析男女的比值比。如需作总的估计, 以用SMR为宜。

二、病例对照分层 $2 \times R$ 表资料的分析: 在病例对照研究中, 如暴露史是按分级收集的, 则其资料可归纳成多个 $2 \times R$ 表(表3)。

表3 病例对照研究资料 $2 \times R$ 归纳表

组别	暴露分级					
	0	1	2	3	4	...
病例	$a_0 (=c)$	a_1	a_2	a_3	a_4	...
对照	$b_0 (=d)$	b_1	b_2	b_3	b_4	...
合计	N_0	N_1	N_2	N_3	N_4	...

根据表中资料可综合分析下列内容:

1. 假设检验: 在 $2 \times R$ 表资料的分析中, 我们着重分析不同暴露级与疾病的关系, 或者

分析其剂量反应关系。在检验这种总趋势时可用 Mantel-Haenszel 扩展公式(见公式 9)。

$$X_{M-EXT} = \frac{\sum_i \left(\sum_j a_j Y_j - \frac{\sum_j a_j}{\sum_j N_j} \sum_j N_j Y_j \right)}{\left\{ \sum_j \frac{\sum_j a_j \sum_j b_j}{(\sum_j N_j)^2 (\sum_j N_j - 1)} \left[\sum_j N_j \sum_j N_j Y_j^2 - (\sum_j N_j Y_j)^2 \right] \right\}^{1/2}} \quad (\text{公式-9})$$

Y_j 是人为的给定暴露级 j 的值, 例如定为 0、1、3、5 等值, i 指 i 层。

2. 求回归系数: 回归系数 b_y 代表线性趋势的斜率。计算用公式-10。

$$b_y = \frac{\sum_i \left\{ \sum_j a_j Y_j - \frac{\sum_j a_j}{\sum_j N_j} \sum_j N_j Y_j \right\}}{\sum_i \left\{ \sum_j N_j Y_j^2 - \frac{(\sum_j N_j Y_j)^2}{\sum_j N_j} \right\}} \quad (\text{公式-10})$$

3. 计算各标准化率比: 标准化率比 (standardized rate ratio, SRR) 又称标准化相对危险度, 可用于分析是否存在剂量反应关系。如存在这种关系, 随着暴露级的增高, SRR 值亦愈大。求 SRR 时, 一般以不暴露或最低暴露级亦标准。应用公式-11 分别计算 SRR_0 、 SRR_1 、 SRR_2 ... 等值。

$$SRR = (\sum a_i d_i / b_i) / (\sum c_i) \quad (\text{公式-11})$$

实例2: Doll 和 Hill (1954年) 进一步将吸烟史分级来研究肺癌的病因^[4], 其资料见表 4。

应用上述公式计算下列有关估计值:

① 假设检验: 将暴露分级定为 0, 2.5、10 和 20。 $\chi_{M-EXT} = 6.665$, $P = 2.66 \times 10^{-11}$

② 回归系数 $b_y = 0.01415$

③ 计算各 SRR: $SRR_0 = 1.0$, $SRR_1 = 1.7$, $SRR_2 = 4.0$, $SRR_3 = 5.5$ 。以上结果说明不同

表 4 不同性别吸烟量与肺癌关系的研究

组别	每日吸烟量(支)				
	0~	1~	5~	15~	
男性	病例	2	33	250	364
	对照	27	55	293	274
	合计	29	88	543	683
	OR	1.00	8.10	11.52	17.93
女性	病例	19	7	19	15
	对照	32	12	10	6
	合计	51	19	29	21
	OR	1.00	0.98	3.20	4.21
SRR	1.00	1.7	4.0	5.5	

吸烟量有明显差别, 随着暴露级的增加则影响愈大, 有一定剂量反应关系。

三、病例对照 1 : R 配比研究资料的分析: 病例对照 1 : R 配比资料可按表 5 格式归纳。

表 5 1 : R 病例对照配比资料归纳

有暴露史的病例数	有暴露史的对照数				
	0	1	2	K	R
1	f_{10}	f_{11}	f_{12}	... f_{1k}	... f_{1R}
0	f_{00}	f_{01}	f_{02}	... f_{0k}	... f_{0R}

f_{1k} 表示 1 个暴露病例与 K 个暴露对照的组数, f_{0k} 表示 1 个非暴露病例与 K 个暴露对照的组数, 1 个病例与 R 个对照配比成为一组。根据表中资料作下列分析:

1. 假设检验：用Miettinen (1969) 介绍的公式-12来检验假设。

$$\chi^2 = \frac{V - \sum_{m=1}^R \frac{(m)(S_m)}{(R+1)}}{\sqrt{\sum_{m=1}^R \frac{(m)(S_m)(R+1-m)}{(R+1)^2}}}$$

(公式-12)

式中V为有暴露史病例与最少有1个无暴露史对照的组数。Sm为有m个暴露人数的总组数。如R=1, 则成1:1配对。则公式-12可省化如公式-13。

$$\chi^2 = \frac{f_{10} - f_{01}}{\sqrt{f_{10} + f_{01}}} \quad (\text{公式-13})$$

2. 比值比点估计：比值比点估计用Miettinen (1970) 介绍的极大似然估计的公式：

$$\frac{V}{R_{ML}} - \sum_{m=1}^R \frac{S_m}{R_{ML} + \frac{R+1-m}{m}} = 0$$

(公式-14)

式中R_{ML}为1:R配比研究的比值比极大似然估计，R大于2时，需用迭代法计算。

3. 比值比区间估计：仍可用公式-5和公式-6来计算。

实例3 Mack等(1976)报道了外源性雌激素与子宫内膜癌的关系^[5]，设计以1:4配比病例对照研究方法。其研究结果资料见表6。

表6 雌激素与子宫内膜癌关系的1:4配比研究

有用雌激素史病例数	有用雌激素史对照数					合计
	0	1	2	3	4	
1	3	17	16	15	5	56
0	0	4	1	1	1	7

表6中3字表示有1个有用雌激素的病例与4个无使用雌激素史的对照者共有3组。表中4字表示无使用雌激素史的病例与1个有用雌激素对照者共有4组。

以上述公式计算有关估计值：

① 检验假设： $\chi = 5.582$, $P = 2.38 \times 10^{-8}$ 有非常显著差异，说明使用雌激素与子宫内膜癌有关。

② 比值比点估计：比值比极大似然估计 $R_{ML} = 7.955$ 。

③ 比值比区间估计，其90%可信限为4.317~14.657。

定群研究资料的分析

一、定群研究cohort study分层的累积发病率分析：将资料按表7格式归纳。

表7 定群研究分层累积发病资料归纳

分组	发病	不发病	合计	发病率
暴露	a _i	c _i	N _{1i}	a _i /N _{1i}
非暴露	b _i	d _i	N _{0i}	b _i /N _{0i}
合计	M _{1i}	M _{0i}	T _i	M _{1i} /T _i

定群研究分层累积发病率分析与公式：

1. 假设检验：用公式-1计算。

2. 相对危险度点估计：可用公式-15、公式-16或公式-17和公式-18计算。

$$RR_a = \frac{\sum a_i N_{0i} / T_i}{\sum b_i N_{1i} / T_i} \quad (\text{公式-15})$$

$$RR_w = \text{EXP} \left[\frac{\sum w_i \ln(RR_i)}{\sum w_i} \right] \quad (\text{公式-16})$$

式中 $W_i = M_{1i} N_{1i} N_{0i} / T_i^2$

$$\sum \frac{a_i - (RR) R_{0i} N_{1i}}{1 - (RR) R_{0i}} = 0 \quad (\text{公式-17})$$

$$R_{0i} = \frac{a_i + N_{0i} + (RR)(b_i + N_{1i})}{2(RR) T_i}$$

$$\left[(a_i + N_{0i} + (RR)(b_i + N_{1i}))^2 \right]$$

$$- \frac{M_{1i}}{(RR) T_i} \Bigg)^{1/2} \quad (\text{公式-18})$$

3. 相对危险度区间估计：可用公式-5和公式-6计算。

4. 特异危险度点估计：用加权特异危险度作点估计较方便(公式-19)

$$AR_w = \frac{\sum W_i AR_i}{\sum W_i} \quad (\text{公式-19})$$

式中AR_i为各层的AR, $W_i = (N_{1i}N_{0i}T_i) / (M_{0i}M_{1i})$ 。

5. 特异危险度区间估计：用公式-20和公式-21来计算。

$$\underline{AR} = AR(1-Z/X) \quad (\text{公式-20})$$

$$\overline{AR} = AR(1+Z/X) \quad (\text{公式-21})$$

求90%可信限时, $Z = 1.645$, 求95%可信限时, $Z = 1.96$ 。

6. 相对危险度非均匀性检验：用公式-22来检验各层间相对危险度是否一致。

$$\chi^2_{HET} = -2 \sum \left\{ a_i \ln \left[\frac{R_{0i} RR N_{1i}}{a_i} \right] + b_i \ln \left[\frac{R_{0i} N_{0i}}{b_i} \right] + c_i \ln \left[\frac{(1-R_{0i} RR) N_{1i}}{c_i} \right] + d_i \ln \left[\frac{(1-R_{0i}) N_{0i}}{d_i} \right] \right\} \quad (\text{公式-22})$$

实例4：在大学组糖尿病研究计划中，发现服用甲糖宁者死亡率较服用安慰剂者为高，但可能有年龄等因素混杂，经用年龄分层后，认为甲糖宁与死亡率的联系未达显著水平，联系的强度亦不够大，资料与计算结果如下（见表8、9）

表8 不同年龄甲糖宁与安慰剂和死亡关系的比较

组别	<55岁		≥55岁	
	死亡数	共人数	死亡数	共人数
甲糖宁	8	106	22	98
安慰剂	5	120	16	85

有关总的估计计算结果如下：

①假设检验： $\chi = 1.092$ $P = 0.138$

②相对危险度点估计： $RR_a = 1.326$,

$RR_w = 1.309$, $RR_{ML} = 1.311$ 。

③ RR_{ML} 区间估计：其90%可信限为0.872

表9 不同治疗方案与死亡率关系的比较

组别	死亡数	共人数	死亡率%
甲糖宁	30	204	14.71
安慰剂	21	205	10.24

~1.972。

④相对危险度非均匀性检验： $\chi^2_{HET} = 0.451$, 自由度 = 1, $P = 0.5$

⑤特异危险度： $AR_w = 0.0343$, 其90%可信限为-0.0174~0.0861。

二、定群研究中分层的人时发病率资料的分析：分层人时发病率资料可按表10形式归纳。

表10 定群研究人时发病率资料的归纳

	病例	人年数	人年发病率
暴露组	a_i	N_{1i}	a_i/N_{1i}
非暴露组	b_i	N_{0i}	b_i/N_{0i}
合计	M_{1i}	T_i	M_{1i}/T_i

用下列有关公式分析：

1. 假设检验：用公式-23。

$$\chi = \frac{\sum a_i - \sum N_{1i} M_{1i} / T_i}{\sqrt{\sum M_{1i} N_{1i} N_{0i} / T_i^2}} \quad (\text{公式-23})$$

2. 相对危险度点估计：用公式-15计算RR, 用公式-24计算 RR_{ML} ,

$$\sum a_i - RR \sum \frac{M_{1i}}{RR + \frac{N_{0i}}{N_{1i}}} = 0 \quad (\text{公式-24})$$

3. 相对危险度区间估计用公式-5和公式-6。

4. 特异危险度点估计用公式-19计算。但 $W_i = (N_{1i}N_{0i}) / M_{1i}$ 。

5. 相对危险度的非均匀性检验：用公式-25来检验各层是否一致。

$$\chi^2_{HET} = -2 \sum \left\{ a_i \ln \left[\frac{RR M_{1i}}{a_i (RR + N_{0i} / N_{1i})} \right] + \right.$$

表11 吸烟者与非吸烟者的冠心病死亡情况

年龄组	吸 烟 者			非 吸 烟 者			RRi
	死亡数	人年数	人年发病率	死亡数	人年数	人年发病率	
35~44	32	52,407	0.00061	2	18,790	0.00011	5.55
45~54	104	43,248	0.00240	12	10,673	0.00112	2.14
55~64	206	28,612	0.00720	28	5,710	0.00490	1.47
65~74	186	12,663	0.01469	28	2,585	0.01083	1.36
75~84	102	5,317	0.01918	31	1,462	0.02120	0.95
合计	630	142,247	0.00443	101	39,220	0.00258	1.72

$$b \ln \left\{ (M_{ii}/b_i) \left(1 - \frac{RR}{RR + N_{0i}/N_{1i}} \right) \right\}$$

(公式-25)

实例5: Doll和Hill (1966)提供的英国男医师吸烟者与未吸烟者不同年龄冠心病死亡资料如表11。

以表11中资料作下列分析:

①假设检验: $\chi = 3.319$, $P = 0.00045$ 总的看来, 两组发病率有显著差异。

②相对危险度点估计: $RR_a = 1.425$

$RR_{ML} = 1.426$ 。

③ RR_{ML} 的90%区间估计为1.699~1.96

④相对危险度非均匀性检验: $\chi^2_{HET} = 12.132$, 自由度为4, $P = 0.016$ 。各层间相对危险度有显著差别, 故不宜用总的相对危险度来估计, 还是以各层资料分别分析为好。

其他一些较复杂的流行病学分析一般都可以编出程序用程序计算器进行分析。相信程序

计算器的推广与应用将提高流行病学分析的水平。

参 考 文 献

1. Rothman KJ, Boice JD. Epidemiologic analysis with a programmable calculator. New ed. Boston: Epidemiology Resources, Inc. 1982:1-44.
2. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959; 22:719.
3. Miettinen OS. Estimability and estimation in case-referent studies. Am J Epidemiol 1976;103:226.
4. Doll R, Hill AB. The mortality [of doctors in relation to their smoking habits. Brit Med J 1954;1:1451.
5. Mack TM, et al. Estrogens and endometrical cancer in a retirement community. New Engl J of Med 1976;294:1262.

《腹泻病专辑》(II) 征订启事

“腹泻病专题委员会”和本刊再度联合编辑出版《腹泻病专辑》(I)。

该书辑入本刊收到和向有关专家征集的论著、技术方法、综述, 还有全国人兽腹泻病病原学学术讨论会的部分材料, 共计六十余篇。全书约20万字, 售价2.0元(邮费在内), 预计于1987年3~4月出版, 一律从邮局汇款: 北京昌平流字五号 中华流行病学杂志编辑部。注明“购腹泻病专辑(II)”, 征订截至1987年4月底止。过期不保证供书。

中华流行病学杂志编辑部