

## 一个应用于医学生物学计数资料 统计分析的新方法

陈冬娥

(湖北医学院卫生学教研室, 武汉)

董长垣

(中国科学院武汉病毒研究所, 武汉)

### A NEW STATISTIC METHOD USED TO ANALYSIS SERIAL COUNT DATA IN MEDICAL AND BIOLOGICAL EXPERIMENT

Chen Dong-e

(Dept. of Hygiene, Hubei Medical College, Wuhan)

Dong Chang-yuan

(Wuhan Institute of Virology, Academia Sinica, Wuhan)

#### 摘 要

$\chi^2$  检验是处理医学、生物学实验中计数资料的经典方法。但是, 由于科学研究向纵深发展, 该法已暴露出了本身的弱点。它对日益增多的有序联列计数资料的统计分析显得无能为力。

本文将日本田口玄一博士的数理理论引入了医学生物学研究领域。把累积法, 权重法与方差分析结合起来, 建立了AVWC法, 并应用于病毒研究的有序联列计数资料的统计分析, 揭示了这类研究数据的内在客观规律, 反应了该法的科学性和实用性。

该论文介绍了AVWC法的依据, 应用及计算过程。作者认为该法是一个较为简单, 精确、实用的统计方法。

本稿1986年2月1日收到

本论文受到湖北医学院卫生学教研室王圣基主任的深切关心和指导, 在此表示感谢!

随着电子计算机等新技术在医学、生物学领域中的应用, 生物科学愈来愈向数量化方面发展。并且科学研究数据的处理, 科学信息分析, 研究论文的描述及总结, 均不再是绝对指标, 而是用相对指标表示。现今已证明, 在揭示生物界规律时, 单因素的分析、研究有很大的局限性; 多因素, 多水平的观察、研究则日臻频繁, 统计方法亦较多。有关定量资料的多因素统计方法既多, 又较灵敏。方差分析(analysis of variance)是其主要统计方法之一<sup>[1]</sup>。该法在发掘科学信息中起着很重要的作用。然而, 多因素计数资料迄今仍多用列联表 $\chi^2$ 检验( $\chi^2$ -test)。由于 $\chi^2$ 检验适用条件的局限性, 致使某些较先进的生物学实验在进行计数资料的整理和分析中, 却不能揭示出内在的科学规律。

本文将以一个肿瘤病毒病因研究课题<sup>[2]</sup>为例, 介绍一个有序联列表计数资料的新处理方法。该法引进了日本田口玄一博士的工业实验设计数理理论<sup>[3-4]</sup>, 将累积法(cumulation method), 权重法(weight method)与方差分析(analysis of variance)结合起来, 对生物学实验资料进行分析, 用以提高资料的统计处理效能。我们将该法称为AVWC方法。

## 一. 该法的一般计算步骤

该法的一般计算步骤如下:

### 一. 累积法

1. 等级累积:  $I' = I$ ;  $II' = I + II$ ;  $III' = I + II + III$ .
2. 求各等级的方差, [ $S_{a(i)}$ , 等级变异]

$$S_{a(i)} = \sum \frac{A^2}{n_R} - \frac{(\sum A)^2}{N}$$

$i$ —— $I'$ ,  $II'$ ,  $\dots$ ,  $k'$ 等级。

$A$ ——各等级的频数。

$n_R$ ——该等级对应横行的合计数。

$\Sigma$ ——(sigma) 求和号。

3. 求各级为1, 其余为0(1, 0资料) $\dots$ ,  $k-1$ 的方差[误差变异,  $S_{e(i)}$ .]

$$S_{e(i)} = \Sigma \left( 1^2 \cdot A + 0^2 \cdot B - \frac{A^2}{n_R} \right)$$

$i$ —— $I'$ ,  $II'$ ,  $\dots$ ,  $k$ 等级。

$A$ ——本等级的频数。

$B$ ——其余等级的频数。

$n_R$ —— $A$ 等级所在行的合计数。

4. 求各等级总方差[总变异,  $S_{t(i)}$ .]

$$S_{i(i)} = A_c - \frac{(A_c)^2}{N}$$

$i$ ——I', II', III', ..... $k$ 等级。

$A_c$ ——各等级所在列的合计频数。

$N$ ——两组的总观察例数。

## 二、权重法 (W) :

对各等级权重。

$$W_i = \frac{1}{p(1-p)}$$

$i$ ——I', II', ..... $k-1$ 等级。

$p$ ——该等级的合计频数与总例数之比值。

三、求累积法与权重法的合并方差[合并等级变异]: 即计算各等级累积方差与同等级权重值的积和。

$$(1) S_A = S_{a(i)} \cdot W_{(i)} + \dots + S_{e(k-1)} \cdot W_{(k-1)}$$

$S_A$ ——各等级方差与同等级权重的积和。

$i$ ——I', II', ..... $k$ 等级。

$$(2) S_B = S_{e(i)} \cdot W_{(i)} + \dots + S_{e(k-1)} \cdot W_{(k-1)}$$

$S_B$ ——各等级误差方差与同等级权重的积和。

$i$ ——I', II', ..... $K$ 等级。

$$(3) S_T = S_{t(i)} \cdot W_{(i)} + \dots + S_{t(k-1)} \cdot W_{(k-1)}$$

$S_T$ ——各等级总方差与同等级权重的积和。

## 四、求F值, 定P值。

### 二、实例计算

某研究<sup>[2]</sup>用单纯疱疹病毒 II 型 (HSV-II) 诱发小白鼠宫颈癌及癌前变与 HSV-II 加上黄芩花 (WC) 或 HSV-II 加桐油提取物 (HHPA) 协同诱发小白鼠宫颈癌及癌前变的作用, 以研究 WC 或 HHPA 与 HSV-II 的协同致癌作用, 实验结果如下表:

表1. 每组诱癌之结果  
Table 1. The Result of Cancer Induced in Each Group

组 别	癌发生数	癌前变数	未发生数	合 计
HSV-I	2	2	50	54
HSV-I+WC	6	6	45	57
HSV-I+HHPA	8	17	32	57
合 计	16	25	127	168

此资料为有序联列计数资料。我们先将 HSV-Ⅱ 组与 HSV-Ⅱ + WC 组 进行两组 差异显著性检验。再作 HSV-Ⅱ 与 HSV-Ⅱ + HHPA 组的差异显著性检验。其计算过程如下：

### 一、累积法：

1. 等级累积：以癌发生数为等级Ⅰ，癌前变频数为等级Ⅱ，以癌未发生数为等级Ⅲ。

HSV-Ⅱ组：Ⅰ' = Ⅰ；Ⅱ' = Ⅰ + Ⅱ；Ⅲ' = Ⅰ + Ⅱ + Ⅲ。

HSV-Ⅱ + WC组：Ⅰ' = Ⅰ；Ⅱ' = Ⅰ + Ⅱ；Ⅲ' = Ⅰ + Ⅱ + Ⅲ。

表2. HSV-Ⅰ组和HSV-Ⅰ+WC组的累积结果  
Table 2. The Result Cumulated HSV-Ⅰ and HSV-Ⅰ+WC Groups

组 别	Ⅰ'	Ⅱ'	Ⅲ'
HSV-Ⅰ	2	4	54
HSV-Ⅰ+WC	6	12	57
合 计	8	16	111

2. 求各等级方差[ $S_{a(i)}$ ，等级变异]

$$S_{a(i)} = \sum \frac{A^2}{n_R} - \frac{(\sum A)^2}{N}$$

$$S_{a(I')} = \frac{2^2}{54} + \frac{6^2}{57} - \frac{8^2}{111} = 0.1291$$

$$S_{a(II')} = \frac{4^2}{54} + \frac{12^2}{57} - \frac{16^2}{111} = 0.5163$$

3. 求两组 1, 0 等级资料的方差[ $S_{e(i)}$ ，误差变异]。

$$S_{e(i)} = \sum (1^2 \cdot A + 0^2 \cdot B - \frac{A^2}{n_R})$$

$$S_{e(I)} = (1^2 \times 2 + 0^2 \times 52 - \frac{2^2}{54}) + (1^2 \times 6 + 0^2 \times 51 - \frac{6^2}{57})$$

$$= (2 - \frac{2^2}{54}) + (6 - \frac{6^2}{57})$$

$$= 7.29$$

其余各等级余此类推。

4. 求各等级总方差[总变异， $S_{t(i)}$ ]

$$S_t(I) = A_c - \frac{(A_c)^2}{N}$$

$$= 8 - \frac{8^2}{111}$$

$$= 7.42$$

其余各等级类推。

## 二、权重法(W): 对各等级权重。

$$W_{(i)} = \frac{1}{p(1-p)}$$

$$W(I) = \frac{1}{\frac{8}{111}(1 - \frac{8}{111})}$$

$$= \frac{111^2}{8(111-8)}$$

$$= 14.95$$

其余各等级以此类推。

## 三、求累积法与权重法的合并方差[合并等级变异]: 即计算各等级累积方差与同等级权重值的积和。

$$S_A = S_a(I') \cdot W(I') + S_a(II') \cdot W(II') + \dots + S_a(k-1) \cdot W(k-1)$$

$$= 0.1291 \times 14.95 + 0.5163 \times 8.1059 = 6.12$$

$$S_E = S_e(I') \cdot W(I') + S_e(II') \cdot W(II') + \dots + S_e(k-1) \cdot W(k-1)$$

$$= 7.29 \times 14.95 + 13.18 \times 8.1059 = 215.88$$

$$S_T = S_t(I') \cdot W(I') + S_t(II') \cdot W(II') + \dots + S_t(k-1) \cdot W(k-1)$$

$$= 7.42 \times 14.95 + 13.69 \times 8.1059$$

$$= 222.00$$

## 四、求F值, 定P值。

表3. 方差分析表  
Table 3. The Table of Variance Analysis

变 异 来 源	自 由 度 (n)	方 差 (SS)	均 方 (MS)	F	P
总变异	220	222			
等级变异	2	6.12	3.06	3.09	< 0.05
误差变异	218	215.88	0.99		

结果表明 HSV-II 与 HSV-II + WC 组比较, 差别有显著性 (P < 0.05)。

HSV-II + HHPA 组与 HSV-II 组仍用上法 ANOVA 检验 F = 5.4649, P < 0.01, 差别有高度显著性。故可以考虑黄芩花 (WC) 和桐油提取物 (HHPA) 都与 HSV-II 有协同致癌作用。

## 三、 讨论和结论

在医学、生物学研究中, 计量资料的统计分析方法较多, 较灵敏<sup>[1]</sup>。但是, 计数资

料,特别是多因素计数资料的统计分析,主要还是经典的 $\chi^2$ 检验<sup>[1]</sup>。众所周知,由于医学生物学科研向纵深发展,这类资料与日俱增。然而,在实际应用中, $\chi^2$ 检验远不能解决这类资料的统计分析。因为 $\chi^2$ 检验不适于分析有序联列计数资料,这是一<sup>[3]</sup>;其二,多维表受理论值影响很大,如1/5以上格子理论值小于5时,则不能应用该法进行检验<sup>[5]</sup>;其三,有些实验即或是在严密的实验设计下进行严密的观察,并且在分析时,无论是否经 Yates 校正,其统计效能均差<sup>[3]</sup>。是故,科学信息不能充分挖掘出来。

本文中所分析的肿瘤病毒病因资料<sup>[2]</sup>,是一个有序联列计数资料,我们采用了将此资料合并,变为0,1资料,再分别对两组用 $\chi^2$ 校正公式检验,其差别无显著性。后我们又用直接概率法计算,结果 $P>0.05$ ,该资料差别还是无显著性。于是,我们又采用了第三个方法,即反正弦函数变量转换,再做方差分析,结果仍无显著性。但是,流行病学资料却表明桐油树含有激活EB病毒的物质<sup>[6]</sup>。某些实验研究又进一步证明了黄芩花具有诱发EB病毒早期抗原和促进EB病毒对B淋巴细胞转化作用<sup>[7]</sup>;桐油提取物和黄芩花均具有促癌作用<sup>[8,9,10]</sup>。同时,HSV-II与WC,HSV-II与HHPA的协同致癌作用之实验研究数据具有直观的协同倾向(表1)。由此提示我们,用以上的统计方法处理该资料的效能不佳,不能灵敏地反应内在的客观规律。为此,我们对这个资料的特性进行了认真分析,大胆地将日本田口玄一博士的数理理论<sup>[3,4]</sup>引入了该研究,进行统计分析,将累积法,权重法与方差分析结合起来,建立了AVWC方法,并对该资料进行了统计处理,得出了科学结论。结果提示:黄芩花和桐油提取物与单纯疱疹II型病毒有协同致癌作用。

我们的实践还证明,按田口玄一在汽车工业设计上仅用累积法与方差分析相结合,或者仅用权重法与方差分析相结合,应用于生物学资料统计分析,科学信息同样挖掘不出来。只有把累积法,权重法与方差分析融为一体,并满足方差分析的正态性,齐同性和可加性,即可得出计数资料的科学结论。我们认为:AVWC法不失为一个较为简便,精确、实用的方法。同时,本文详细地介绍了该法的计算方法和过程。

### 参 考 文 献

- [1] R.A.Fisher, 1958, *Statistical Methods for Research Workers*, 13th Ed. Printed and Published in Great Britain by Oliver and Boyd Ltd., Edinburgh,
- [2] 孙瑜等人, 医学微生物学杂志, 待发表。
- [3] 田口玄一, 1976, 实验计画法(第三版), 上册, 丸善株式会社。
- [4] 田口玄一, 1977, 实验计画法(第三版), 下册, 丸善株式会社。
- [5] 上海第一医学院卫生统计学教研室, 1979, 医学统计方法, 上海科学出版社。
- [6] Zeng Y. et al., 1984, *Cancer Letter*, 23:53.
- [7] 芩花根协作组, 1978, 医药工业, 1:6.
- [8] Ito y. et al., 1983, *The 4th International Symposium of the Princess Takamatsu Cancer Research Fund*, Tokyo, Japan.
- [9] 胡银玲等, 待发表。
- [10] 孙瑜等人, 1985, 湖北医学院学报, 6(3):213.