第4卷第2期　　福建省农科院学报　　1989年11月

r×s列联表的对数线性模型在农业试验分析上的应用

郭如灼　　张开泰

（福建省农学院基础部、园艺系）

提　　要

列联表的对数线性模型具有拟合程度高，对期望频数的下限不提任何要求，计算特别简便的特点。本文介绍二维r×s列联表的一般形式及其对数线性模型，并利用2×s列联表的对数线性模型分析四种植物生长调节剂在荔枝雌花期处理对座果率的影响。

二维列联表起始于Maxwell的"定性数据分析"，Everitt等把它推广到多维方面，Ku，H.H.等提出用对数线性拟合模型对多元计数数据作全面精致的分析[1,2]。列联表的对数线性模型不仅拟合程度高，在二维、三维的情况下，只要应用带有对数的计算器就能进行计算，而且对期望频数的下限不必限制（非对数线性模型一般要求期望频数$N\cdot P_{i\cdot}\cdot P_{\cdot j}>5$），自由度不小于1就可以。因此它是一种行之有效而且简便、精确的方法。此法在医学上应的用已有许多报道[1,3]，在农业试验中的计数数据分析上也是有用的。本文利用二维2×s列联表的对数线性模型来分析荔枝雌花期用植物生长调节剂处理后对座果率的影响，作为探讨提高荔枝产量的一个途径。

一　二维r×s列联表的独立性检验

（一）　二维r×s列联表的一般形式　设原因因素B有S个水平（B_1，B_2，……B_s），B的每个水平对研究对象A的r个不同结果（A_1，A_2，……A_r）都可能有影响，这样就产生了二维r×s列联表，其一般形式如表1。

表中n_{ij}为原因B取j（j=1，2,……,s）水平时，研究对象A的第i（i=1，2,……，r）种结果发生的频数[1] $n_{\cdot j}=\sum_{i=1}^{r}n_{ij}$是B取第j水平时各种结果发生的频数总和，即B的第j水平试验观察总数（即第j列累加），$n_{i\cdot}=\sum_{j=1}^{s}n_{ij}$是B取所有水平时第i种结果

本文于1987年6月20日收到。

发生频数的总和(即第i行的累加)，$N=\sum_{j=1}^{s}n_{j.}=\sum_{i=1}^{r}n_{i.}$

表1　　二维r×s列联表的一般形式

结果(A)	原因(B) B₁	B₂	…	B_j	…	B_s	$n_{i.}$
A_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1s}	$n_{1.}$
A_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2s}	$n_{2.}$
⋮	⋮	⋮	…	⋮	…	⋮	⋮
A_i	n_{r1}	n_{i2}	…	n_{ij}	…	n_{is}	$n_{i.}$
⋮	⋮	⋮	…	⋮	…	⋮	⋮
A_r	n_{r1}	n_{r2}	…	n_{jr}	…	n_{rs}	$n_{r.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	…	$n_{.j}$	…	$n_{.s}$	N

（二）　检验目的　为了检验原因B的各水平对结果A是否独立（即是否有显著的影响）；或B的某些水平对A是否独立。

（三）　检验列联表独立性的对数线性拟合模型[2]设

$$\chi^2=2\left(N\ln N+\sum_{i=1}^{r}\sum_{j=1}^{s}n_{ij}\ln n_{ij}-\sum_{j=1}^{s}n_{.j}\ln n_{.j}-\sum_{i=1}^{r}n_{i.}\ln n_{i.}\right)\quad\cdots\cdots(1)$$

其自由度$df=(r-1)(s-1)$。定义$0\ln 0=\lim\limits_{x\to 0}x\ln x=0$。

指定显著性水平α，则当$\chi^2<\chi^2_\alpha[(r-1)(s-1)]$时，B的各水平对A是独立的(即在显著性水平α下B的各水平对A的影响无显著差异)；若$\chi^2\geqslant\chi^2_\alpha[(r-1)(s-1)]$，则B的各水平对A不独立(即是连带的，或说在显著性水平α下B的各水平对A的影响有显著的差异)。其中$\chi^2_\alpha[(r-1)(s-1)]$从$\chi^2$（卡方）分布表查得。

（四）　把2×S列联表分解为非独的2×2表检验　当B的S个水平中有一个是对照的水平时，我们感兴趣的是B的其他S－1个水平中的每一水平与对照的水平比较对A是否独立。这样要将一张2×S列联表分解成S－1张2×2列联表，然而这是非独立的分解，把它们当作自由度为1的x^2变量进行检验是不妥当的。Brunden于1972年指出，只要选定显著性水平α，然后将从S－1张2×2表得到的每一χ^2值与水平

$$\alpha=\alpha'/2(s-1)\quad\cdots\cdots(2)$$

下自由度为1的x^2表查得值作比较即可[1]。

二　用二维2×S列联表分析研究不同生长调节剂在荔枝雌花期处理对座果率的影响

我们用四种不同的生长调节剂各取三个水平在荔枝雌花期进行全树冠喷洒，分别于

处理后20天、40天观察其座果数与落果数。现将观察的结果列于表2。

表2 不同生长调节剂荔枝雌花期处理后的观察值

处理 \ 观察值 \ 调节剂		萘乙酸 (ppm)			九二〇 (ppm)			卅烷醇 (ppm)			2,4-D (ppm)			清水对照	$n_{i.}$
		5	10	15	25	50	100	1.0	1.5	2.0	5	10	15		
20天后	座果数	232	409	340	371	403	387	135	132	155	224	307	313	241	3647
	落果数	1963	1678	1967	1718	1918	1715	2070	1968	2051	1792	1816	1744	2079	24481
	$n_{.j}$	2195	2087	2307	2089	2321	2102	2205	2100	2206	2016	2123	2057	2320	28128
	平均座果率	0.106	0.196	0.147	0.178	0.174	0.184	0.061	0.063	0.070	0.111	0.145	0.152	0.104	0.130
40天后	座果数	166	180	171	170	182	187	80	82	95	127	150	145	114	1849
	落果数	2029	1907	2136	1919	2139	1915	2125	2018	2111	1889	1973	1912	2206	26279
	$n_{.j}$	2157	2087	2307	2089	2321	2125	2205	2100	2206	2016	2123	2057	2320	28128
	平均座果率	0.076	0.086	0.074	0.081	0.087	0.089	0.036	0.039	0.043	0.063	0.071	0.071	0.049	0.070

表2中$n_{.j}$就是B取第j水平时雌花试验观察的总数。

（一） 检验不同生长调节剂的各水平及对照水平对座果率的影响是否独立 对表2中的20天观察数据按公式（1）计算得

$$\chi^2 = 2\,[\,28128\ln 28128 + (232\ln 232 + 409\ln 409 + \cdots + 313\ln 313 + 241\ln 241 + 1963\ln 1963 + 1678\ln 1678 + \cdots + 1744\ln 1744 + 2079\ln 2079) - (2195\ln 2195 + 2087\ln 2087 + \cdots + 2057\ln 2057 + 2320\ln 2320) - (3647\ln 3647 + 24481\ln 24481)\,] = 541.32。$$

同样方法从表2中的40天的观察数据计算得

$$\chi^2 = 148.349。$$

查 $df = (2-1)(13-1) = 12$ 的 χ^2 分布表得 $\chi_{0.001}(12) = 34.54$。

可见不同生长调节剂的不同水平及对照水平，对荔枝雌花期处理后的20天、40天的座果率的影响，有极为显著的差异。因此下面还有进一步分析的必要。

（二） 检验每一种生长调节剂的各水平及对照水平对座果率是否独立 每种生长调节剂有三个水平，加上对照水平就构成 2×4 列联表。我们从萘乙酸处理后 20 天的观察值为例进行计算，所用的 2×4 列联表见表3。

表3 萘乙酸不同浓度荔枝雌花期处理20天后及对照组的观察值

浓度(ppm)	5	10	15	清水	$n_{i.}$
座果数	232	409	340	241	1222
落果数	1963	1678	1967	2097	7687
$n_{.j}$	2195	2087	2307	2320	8909

按公式（1）计算得

$\chi^2 = 2[\,8909\ln 8909 + (232\ln 232 + 409\ln 409 + 340\ln 340 + 241\ln 241 + 1963\ln 1963 + 1678\ln 1678 + 1967\ln 1967 + 2079\ln 2079) - (2195\ln 2195 + 2087\ln 2087 + 2307\ln 2307 + 2320\ln 2320) - 1222\ln 1222 + 7687\ln 7687\,) = 255.615$

同样方法计算其他3张及处理后40天观察值的4张2×4表的χ^2值。现将计算的结果列于表4。

表4 不同生长调节剂的不同水平及对照水平构成列联表的值

调节剂		萘乙酸	九二〇	卅烷醇	2,4-D
χ^2值	20天	255.62***	77.18***	-36.956***	33.44***
	40天	26.66***	32.28***	-5.194	12.05**

$\chi^2_{0.05}(3)=7.815$　　$\chi^2_{0.01}(3)=11.345$　　$\chi^2_{0.005}(3)=12.838$

注：表中χ^2值的前面带负号的表示该处理中各水平使平均座果率低于对照水平的平均座果率（下同）。

从表4中χ^2_L值的大小得知，处理后20天时，每一种生长调节剂浓度的变化（含清水对照）对座果率的影响都有极为显著的差异。差异最大的是萘乙酸，其次是九二〇，再次是卅烷醇，最后是2,4-D。

40天后检查，除卅烷醇外，其他三种生长调节剂浓度的变化（含清水对照）对座果率的影响都有极为显著的差异。差异最大的是九二〇，其次是萘乙酸，再次是2,4-D。卅烷醇浓度的变化（含清水对照）对座果率的影响无显著的差异。

（三）检验每一种生长调节剂的各水平对座果率是否独立　我们再以萘乙酸处理后20天的观察值为例进行计算，所用的2×3表见表5。

表5 萘乙酸不同水平在荔枝雌花期处理20天时观察值

浓度(ppm)	5	10	15	$n_{i.}$
座果数	232	409	340	981
落果数	1963	1678	1967	5608
$n_{.j}$	2195	2087	2307	6589

按公式（1）计算得

$\chi^2 = 2[6589\ln 6589 + (232\ln 232 + 409\ln 409 + 340\ln 340 + 1963\ln 1963 + 1678\ln 1678 + 1967\ln 1967) - (2195\ln 2195 + 2087\ln 2087 + 2307\ln 2307) - (981\ln 981 + 5608\ln 5608)] = 69.187$。

类似地计算其他3张2×3表及处理后40天观察值的4张2×3表，现将计算结果列于表6。

表6　不同生长调节剂的不同水平构成的列联表的 χ^2 值

调节剂		萘乙酸	九二0	卅烷醇	2,4-D
χ^2 值	20天	69.187***	0.834	-1.664	8.487*
	40天	2.562	1.682	-1.348	1.246

$\chi^2_{0.05}(2)=5.991$　　$\chi^2_{0.01}(2)=9.210$　　$\chi^2_{0.005}(2)=10.597$

从表6中的 χ^2 值得知，生长调节剂处理后20天检查，萘乙酸的各水平对座果率的影响有极为显著的差异，2,4-D的各水平对座果率的影响有显著的差异，而九二〇及卅烷醇的各自三个水平对座果率的影响是独立的。40天后各种生长调节剂的各自水三个平对座果率的影响都是独立的。因此只有对处理后20天的观察值中萘乙酸及2,4-D的各水平对座果率的影响还有进一步分析的必要，但为了彻底讨论，下面仍将每一种处理（共12种处理）都与对照组作比较。

（四）　检验各种生长调节剂的每一水平与对照组对座果率的影响是否独立　我们以萘乙酸5ppm处理后20天与对照组的观察值为例进行计算，所用的2×2表见表7。

表7　萘乙酸5ppm处理后20天与对照组观察值

处理	萘乙酸5ppm	对照	$n_{i.}$
座果数	232	241	473
落果数	1963	2079	4042
$n_{.j}$	2195	2320	4515

按公式（1）计算得

$\chi^2=2〔4515\ln4515+(232\ln232+241\ln241+1963\ln1963+2079\ln2079)-(2195\ln 2195+2320\ln2320)-(473\ln473+4042\ln4042)〕=0.0396\approx0.040$

类似地可以计算其他11张2×2表及处理后40天观察值的12张×2表。现将计算结果列于表8。

表7是从表3分解出来的三张2×2表之一，这是非独立的分解，必须将指定的显著性水平α除以2（4－1）＝6（按公式（2）），然后作检验。例如取α＝0.05，则应在显著性水平

$\alpha'=0.05\div6=0.08\dot{3}\approx0.01$ 下作检验。

从表8可知，生长调节剂处理后20天时除萘乙酸5ppm及2,4-D5ppm外，其余10个处理都使座果率与对照组比较有极为显著的提高。处理后40天，除2,4-D5ppm及卅烷醇的各水平外，其余各种处理都使座果率比对照组有显著或极为显著的提高，卅烷醇的各水平使座果率比对照组低，但未达到显著的地步。

三　结束语

列联表的对数线性模型用于独立性检验，除本文开头所指出的优点外，它比假设检验还具有如下的优点：

表8　　各种生长调节剂的每一水平与对照水平观察值组成的列联表

调节剂 (ppm)		萘乙酸			九二〇		
		5	10	15	25	50	100
χ^2 值	20天	0.040	74.476***	20.016***	50.092***	47.697***	58.486***
	40天	13.652**	24.374***	12.569**	18.991**	16.795**	27.696***

$\chi^2_{0.05}(1)=3.84$，$\chi^2_{0.001}(1)=6.635$，

卅烷醇			2,4-D		
1.0	1.5	2.0	5	10	15
-27.389***	-24.427***	-16.145**	0.589	17.002**	22.96***
-4.582	-2.663	-0.949	3.934	9.184*	8.915*

$\chi^2_{0.005}(1)=7.879$，$\alpha'=0.05\div6\approx0.01$

（一）　可以对各因素影响的多种（r种）结果进行差异显著性的检验，而假设检验只能对各因素影响的某一种结果进行检验。

（二）　计算时涉及的数据少于假设检验。对本试验来说，每个水平有40次重复，对每两个水平作差异显著性的检验时，假设检验需用80个数据，而同样的检验用列联表分析时用的是2×2表，只涉及到9个数据，而且计算手续非常简便，不易出差错。

（三）　作假设检验时，B的每个水平都需要重复试验，重复的次数不大时精确度就不高，而作列联表分析时B的每个水平只需作一次试验。但不论作那一种检验样本容量都尽可能大些。

（四）　表8中χ^2值的大小顺序恰好是表2中相应的平均座果率大小的顺序，这是假设检验所不具有的，也说明列联表分析的对数线性模型检验更为精确。

然而列联表分析只适用于对各个结果能统计其发生频数的各种结果发生率的显著性检验，因而在应用上不如假设检验广泛，这是它的局限性。

本文在实例中提出四种不同的列联表进行分析，应用时可根据需要取其一种、部分或全部。

参　考　文　献

〔1〕　B.S.艾沃日特，1980，列联表分析，科学出版社，北京

〔2〕　汤旦林，1982，生物统计原理与多元分析方法。科研中的统计方法，1982增刊1，生物统计(杂志）22：74～95

〔3〕　Ku, H.H. and S.kullback, 1974 Log-linear models in contingency tabel analysis, Amer. statistician, 28; 115～122

THE UTILIZATION OF THE LOG-LINEAR MODEL OF THE R×S CONTINGENCY TABLE IN AGRICULTURAL EXPERIMENT ANALYSIS

Guo Ruzhuo, Zhang Kaitai

(Basic Parts and Horticultural Dept. Fujian Agricultural College)

ABSTRACT

This paper presented the log-linear model of the r×s contingency table to be used on the analysis of the effects of plant growth regulators which were used in the female flowers blooming of cv. Lantek Litchi. The crowns were sprayed with GA_3, NAA, 2,4-D and tri acontanol for raising fruit setting percentage. When 20 days after treatments the results showed very significant differences among the levels of NAA and 2,4-D, but hadn't shown difference among the levels of GA_3 and triacontanol. Compared with CK, GA_3, NAA and 2,4-D came up to very significant difference for raising fruit setting percentage except NAA 5 ppm and 2,4-D 5 ppm, and triacontanol came up to very significant negative levels. When 40 days after treatments, no significant difference among the levels of each regulators were observed. The results compared with CK had shown NAA and GA_3 came up to very significant levels for raising fruit setting percentage, 2,4—D came up to significant levels except 5 ppm, but triacontanol reduced fruit setting percentage and didn't come up to significant levels.