高维列联表

高维列联表的数据结构

三维r×c×t列联表的数据结构：

假设n个个体按照三个属性分类，其中属性A有r类，属性B有c类，属性C有t类;

n个个体中属于A_i、B_j、C_k类的有n_{ijk}个，联合概率为p_{ijk}

C	A	B		
		B_1	$\cdots$	B_c
C_1	A_1	p_{111}	$\cdots$	p_{1c1}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	A_r	p_{r11}	$\cdots$	p_{rc1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_t	A_1	p_{11t}	$\cdots$	p_{1ct}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	A_r	p_{r1t}	$\cdots$	p_{rct}

高维列联表的结构

【例1】为了解不同年龄的男性，吸烟与呼吸系统疾病之间的关系，调查数据见下表：

		不吸烟	吸烟	合计	χ^2 (P 值)
年龄< 40	呼吸正常	567	874	1441	0.620
	呼吸不正常	14	28	42	(0.431)
年龄40-59	呼吸正常	328	780	1108	23.355
	呼吸不正常	2	68	70	(0.001)

上表为三维2×2×2列联表。其中，“年龄”为层属性，“呼吸情况”为行属性，“吸烟情况”为列属性。在每一层，都是一个二维列联表。

高维列联表的压缩

通过把不同年龄的数据合并，可以将三维列联表压缩成二维列联表。

表1

	不吸烟	吸烟	合计	χ^2 (P 值)
呼吸正常	895	1654	2549	20.668
呼吸不正常	16	96	112	(0.000)
合计	911	1750	2661	

也可以合并“呼吸情况”的数据，得到“年龄与吸烟情况”的二维表；

或合并“吸烟情况”的数据，得到“年龄与呼吸情况”的二维表。

高维列联表的压缩

- 一般地，列联表压缩后的数据结构为：

	B_1	---	B_c	合计
A_1 ⋮ A_r	n_{11+} ⋮ n_{r1+}	$\cdots$ $\ddots$ $\cdots$	n_{1c+} ⋮ n_{rc+}	n_{1++} ⋮ n_{r++}
合计	n_{+1+}	$\cdots$	n_{+c+}	n

- 经过合并压缩后得到的二维表，称为边缘表。
- 边缘表实际上是“忽略”某个属性后得到的列联表。

高维列联表的分层

✧与压缩相反，可以把三维表中的**每一层的二维表**分离出来加以研究，这时称为**部分表**。

✧上述三维2×2×2列联表，可以通过按年龄分层，分离出两张二维列联表，即两个部分表。

✧**部分表中的关联性称为条件关联性**，即某个属性给定(被控制)时，另外两个属性之间的关系。

✧部分表的**条件关联性**可能和**边缘表中的关联性**有较大差异，甚至是**自相矛盾（辛普森悖论）**。

✧正是边缘表与部分表分析的条件发生变化，所以把**压缩与分层结合**起来分析是完全必要的。

部分表与边缘表

从分层后的两张二维表（部分表）中，根据各自的卡方值可以看出，

在年龄<40的部分表中，吸烟情况与呼吸情况是相互独立的；

而在另一个部分表即年龄40~59的二维表中，二者则是相关联的（或不独立）。

从按年龄合并、压缩后的二维表（边缘表）来看，吸烟情况与呼吸情况之间是相关联的。

可见，部分表与边缘表关联性不一致，有时甚至会明显矛盾，完全相反。

部分表与边缘表

从四格表可知，优势比可以用来度量属性之间的关联性；

根据部分表计算的优势比，称为**条件优势比**；

根据边缘表计算的优势比，称为**边缘优势比**；

与前面所述的部分表与边缘表的关系相一致，条件优势比与边缘优势比是不同的，有时二者会给出完全相反的结论；

当部分表中两个属性变量条件独立时，所有的条件优势比都等于1；但根据边缘表计算的边缘优势比可能并不等于1，即条件独立不代表边缘独立。

高维列联表的分层

分层与压缩相类似地，都可以按照不同的属性压缩或者分层。

一般地，按属性A分层，可以分成r个二维c×t列联表；按属性B分层，可以得到c个二维r×t列联表；按属性C分层，可以得到t个二维r×c列联表.

压缩与分层都是针对高维列联表的分析方法，是从不同角度和途径对不同属性之间的关系进行分析的需要。

基于辛普森悖论的存在，压缩与分层经常结合起来使用。

高维列联表的条件独立性检验

对于三维列联表，按照某一属性进行分层研究，实际上就是要进行条件独立性检验。

如：考虑**C给定后A与B条件**独立问题的检验。第k个二维r×c列联表的检验统计量应为:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ijk} - n_{i+k}n_{+jk}/n_{++k})^2}{n_{i+k}n_{+jk}/n_{++k}} \sim \chi^2((r-1)(c-1))$$

$$G^2 = -2\sum_{i=1}^{r}\sum_{j=1}^{c} n_{ijk}\ln\left(\frac{n_{i+k}n_{+jk}}{n_{++k}n_{ijk}}\right) \sim \chi^2((r-1)(c-1))$$

为此，可以把t个统计量的和作为条件独立性检验的检验统计量，其自由度为t(r-1)(c-1)。

高维列联表的条件独立性检验

- 另外还有两个条件独立性检验的问题，即A给定B与C条件独立和B给定A与C条件独立。
- 这三种条件独立性检验见下表：

原假设	期望频数	检验统计量	自由度
C给定A、B独立	$\frac{n_{i+k}n_{+jk}}{n_{++k}}$	$-2\sum\sum\sum n_{ijk}\ln\left(\frac{n_{i+k}n_{+jk}}{n_{++k}n_{ijk}}\right)$	$t(r-1)(c-1)$
A给定B 、C独立	$\frac{n_{ij+}n_{i+k}}{n_{i++}}$	$-2\sum\sum\sum n_{ijk}\ln\left(\frac{n_{ij+}n_{i+k}}{n_{i++}n_{ijk}}\right)$	$r(c-1)(t-1)$
B给定A、C独立	$\frac{n_{ij+}n_{+jk}}{n_{+j+}}$	$-2\sum\sum\sum n_{ijk}\ln\left(\frac{n_{ij+}n_{+jk}}{n_{+j+}n_{ijk}}\right)$	$c(r-1)(t-1)$

表2

专业	性别	录取人数	未录取人数	合计	录取比例
A1	男生	423	227	650	65.1%
	女生	33	17	50	66.0%
A2	男生	162	263	425	38.1%
	女生	200	345	545	36.7%
A3	男生	96	214	310	31.0%
	女生	90	184	274	32.7%
A4	男生	45	136	181	24.9%
	女生	94	296	390	24.1%
A5	男生	7	86	93	7.5%
	女生	11	143	154	7.1%

条件独立性检验举例

【例2】某研究生院某年有1659名男生和1413名女生申请报考，其中有733名男生和428名女生被录取。录取结果见表5.12：

表3

	录取人数	未录取人数	合计	录取比例
男生	733	926	1659	44.2%
女生	428	985	1413	30.3%

- 经检验，该表的G^2=63.18，p值<0.0001，故认为该校有偏爱男生的倾向。
- 为此，学校想进一步了解哪些专业有偏爱男生的倾向，根据5个不同专业的招生录取情况进行分析（见表2）。
  - 例题中的表3是由表2合并压缩而成。从上表可以看出各个专业的男女生录取比例基本上一致，为此需要进行条件独立性检验。

条件独立性检验举例

将表2按专业分层后形成5个部分表，各自独立性检验统计量的值分别为：0.0175，0.2058，0.2364，0.0386，0.0126；

将5个统计量的值加起来才是条件独立性检验统计量的值，应为0.5109，自由度为5；检验的p值=0.9917，故接受原假设，认为性别与是否录取无关，验证了男女录取比例基本一致的判断。

可见，同样的数据合起来（边缘表）与分开来（部分表）的关联性检验的结论正好相反。这就是辛普森悖论。

条件独立性检验举例

辛普森悖论产生的主要原因在于：

在计算总的录取比例时，尽管各个专业的男女生录取比例没有显著差异，但是男生和女生所采用的权重相差较大。

其中，在计算男生录取比例时，录取比例高的专业权重大，录取比例低的专业权重小，导致男生总的录取比例偏高；

在计算女生录取比例时，录取比例高的专业权重小，而录取比例低的专业权重大，从而使总的录取比例偏小。

因此，经过检验，不能说该校有偏爱男生的倾向。

高维列联表的独立性检验

三维列联表除面临前述的条件独立性检验外，还会遇到另外两种独立性检验问题。

A、B、C 相互独立	(A, B, C)	$p_{ijk} = p_{i++}p_{+j+}p_{++k}$
A 和(B, C) 相互独立	(A, BC)	$p_{ijk} = p_{i++}p_{+jk}$
B 和(A, C) 相互独立	(B, AC)	$p_{ijk} = p_{+j+}p_{i+k}$
C 和(A, B) 相互独立	(C, AB)	$p_{ijk} = p_{++k}p_{ij+}$
A 给定后B 和C 条件独立	(AB, AC)	$p_{ijk} = p_{ij+}p_{i+k}/p_{i++}$
B 给定后A 和C 条件独立	(BA, BC)	$p_{ijk} = p_{ij+}p_{+jk}/p_{+j+}$
C 给定后A 和B 条件独立	(CA, CB)	$p_{ijk} = p_{i+k}p_{+jk}/p_{++k}$

高维列联表的独立性检验

以上三种情况下的独立性检验问题之间有以下关系：其中，由左到右是包含和推出的关系，所描述的模型也由简单到复杂。

$$(A,B,C) \Rightarrow \begin{cases} (A,BC) \Rightarrow \begin{cases} (BA,BC) \\ (CA,CB) \end{cases} \\ \\ (B,AC) \Rightarrow \begin{cases} (AB,AC) \\ (CA,CB) \end{cases} \\ \\ (C,AB) \Rightarrow \begin{cases} (AB,AC) \\ (BA,BC) \end{cases} \end{cases}$$

为此，可以在处理三维列联表时，按照以上顺序进行检验。如果前面的检验没有被拒绝，就可以不用再进行后面的检验。

高维列联表的独立性检验

对于第一种情况下，原假设为：

任意格(i,j,k)的期望频数为：$p_{ijk} = p_{i++} p_{+j+} p_{++k}$

其似然比统计量为：$n\hat{p}_{i++} \hat{p}_{+j+} \hat{p}_{++k} = \dfrac{n_{i++} n_{+j+} n_{++k}}{n^2}$

$$G^2 = -2 \ln \Lambda = -2 \sum \sum \sum n_{ijk} \ln \left(\frac{n_{i++} n_{+j+} n_{++k}}{n^2 n_{ijk}} \right)$$

$$\sim \chi^2 (rct - r - c - t + 2)$$

高维列联表的独立性检验

对于第二种情况下，原假设：A和(B、C)相互独立，相当于把后两种属性组成一种新的属性(BC),因此原假设可以记为：

$$p_{ijk} = p_{i++}p_{+jk}$$

任意格上的期望频数应为： $n\hat{p}_{i++}\hat{p}_{+jk} = \dfrac{n_{i++}n_{+jk}}{n}$

其似然比统计量应为：

$$G^2 = -2\ln\Lambda = -2\sum\sum\sum n_{ijk}\ln\left(\frac{n_{i++}n_{+jk}}{nn_{ijk}}\right)$$

$$\sim \chi^2((r-1)(ct-1))$$

高维列联表的独立性检验

- 如果(A,BC)没有被拒绝，这时独立性检验就结束了。接下来可以分析二维B×C列联表；
- 如果(A,BC)被拒绝，需要类似地检验另外两种情形即(B,AC)和(C, AB).其期望频数及统计量见表：

原假设	期望频数	统计量	自由度
(B,AC)	$\frac{n_{+j+}n_{i+k}}{n}$	$-2\sum\sum\sum n_{ijk}\ln\left(\frac{n_{+j+}n_{i+k}}{nn_{ijk}}\right)$	$(c-1)(rt-1)$
(C,AB)	$\frac{n_{++k}n_{ij+}}{n}$	$-2\sum\sum\sum n_{ijk}\ln\left(\frac{n_{++k}n_{ij+}}{nn_{ijk}}\right)$	$(t-1)(rc-1)$

高维列联表的独立性检验

如果第二种情况所有情形都被拒绝时，接下来进行条件独立性检验。

如果某个条件独立性检验没有被拒绝，则独立性讨论结束。接下来可以按照压缩和分层方法进行分析。

如果条件独立性检验都被拒绝时，说明三种属性之间具有相关关系，需要进一步分析，以确定是否仅两两相关，还是包括三次交互效应（称为饱和模型）。

独立性检验举例

【例3】某保险公司某年有12299份保单，有赔款记录的保单和无赔款记录的保单按照车辆类型和被保险人年龄进行统计后结果见表：

表4

		有赔款记录的保单数	无赔款记录的保单数
普通车	年龄< 25	741	2829
	年龄≥ 25	882	4945
高性能车	年龄< 25	453	1169
	年龄≥ 25	248	1032

问：以上三种属性之间有何关系。

独立性检验举例

- 用A表示车型，B表示年龄，C表示有无赔款记录，独立性检验结果如下：

原假设	G^2	P值
(A,B,C)	431.219	$P(\chi^2(4) \geq G^2) \approx 0$
(A,BC) (B,AC) (C, AB)	333.6031 365.5279 142.7808	$P(\chi^2(3) \geq G^2) \approx 0$ $P(\chi^2(3) \geq G^2) \approx 0$ $P(\chi^2(3) \geq G^2) \approx 0$
(AB,AC) (BA,BC) (CA, CB)	77.0895 45.1764 267.9117	$P(\chi^2(2) \geq G^2) \approx 0$ $P(\chi^2(2) \geq G^2) \approx 0$ $P(\chi^2(2) \geq G^2) \approx 0$

独立性检验举例

可见，独立性检验全被拒绝了，说明三个属性间仅有相关关系。为此，对普通车和高性能车分别进行相合性度量和检验。结果如下：

	τ (SE)	γ (SE)	$d_{C\|R}$ (SE)	$d_{R\|C}$ (SE)
普通车	0.0722 (0.0105)	0.1898 (0.0266)	0.0566 (0.0083)	0.0927 (0.00135)
高性能车	0.0992 (0.0181)	0.2345 (0.0424)	0.085 (0.0157)	0.1157 (0.0210)

经分别进行相合检验发现：无论何种车型，都是年龄越小，有赔款记录的可能性越大。

Cochran-Mantel-Haenszel 和Breslow-Day检验

在前面的分析中，仅分别按普通车和高档车进行了相合性检验，这样的分析还不够，需要进一步做条件相合性检验。

即对于r×2×2的三维列联表，在层属性A给定后，行属性B与列属性C是否条件相合（正、负）。

条件相合性检验的原假设、备择假设分别为：

H_0:A给定后B和C条件独立；

H_1:A给定后B和C条件正相合(或负相合)；

每一层的相合检验统计量为：

$$U_i = (n_{i11} - E(n_{i11})) / \sqrt{D(n_{i11})}$$

r×2×2的条件相合性检验

那么，r个统计量U_i合并后，可以综合为一个检验统计量：

$$U=\frac{\sum_{i=1}^{r} n_{i11}-\sum_{i=1}^{r}\frac{n_{i1+}n_{i+1}}{n_{i++}}}{\sqrt{\sum_{i=1}^{r}\frac{n_{i1+}n_{i2+}n_{i+1}n_{i+2}}{n_{i++}^{2}(n_{i++}-1)}}}\sim N(0,1)$$

当U比较大时，认为A给定后B与C条件正相合；
当U比较小时，认为A给定后B与C条件负相合。
在无方向检验时，可以采用U^2作为检验统计量。
以上检验方法称为CMH检验。

r×2×2的条件相合性检验

【例4】计算例3中A给定后B和C条件正相合检验问题的解。

经计算，U=8.7742，认为A给定后B与C条件正相合，即无论什么车型，年龄越小，有赔款记录的可能越大。

需要指出的是，这一检验只有在前面进行了分层分析检验的基础上才有意义。

这是因为，有时候会出现有的层正相合，有的层负相合，最后做条件相合性检验却发现条件独立。

相合程度的检验

在四格表中，当用优比衡量相合性时，θ>1为正相合, θ<1为负相合，θ=1为相互独立。

在三维列联表中，可以用θ_i表示A给定为A_i后B和C的条件相合性的程度。

那么，各层四格表相合程度是否相同的检验问题，原假设和备择假设应为：

$$H_0:\theta_1=\cdots=\theta_r; H_1:\theta_1,\cdots,\theta_r\text{不全相等}$$

检验统计量为：

$$\chi^2=\sum_{i=1}^{r}(\eta_i-\overline{\eta})^2/a_i \sim \chi^2(r-1)\text{，其中}$$

$$\eta_i=\ln\hat{\theta}_i,\quad \overline{\eta}=\frac{\sum(\eta_i/a_i)}{\sum(1/a_i)},\quad a_i=1/n_{i11}+1/n_{i12}+1/n_{i21}+1/n_{i22}$$

相合程度的检验

当卡方值较大时，认为各层的相合程度不相同；而当卡方值较小时，认为各层的相合度相同。这一检验方法称为Breslow-Dayχ^2检验.

【例5】计算例3中A给定后B和C条件相合程度。

经计算， $\chi^2 = 0.7887, P(\chi^2(1) \geq 0.7887) = 0.3754$

故认为各层B和C条件正相合的程度没有显著区别。

把各层相合程度相同称为齐性，或齐次关联性；

齐性具有对称性。在控制任意的第三变量的情况下，剩余两个变量的条件优势比都相同，或称这两个变量对第三个变量没有交互作用。

例6

性别	年龄	对某项措施看法		同意比例	Sommers d 系数
		不同意	同意		
男	年龄<40	90	110	55%	d(c|r)=0.150
	年龄40~50	540	1260	70%	标准差 0.037
女	年龄<40	270	1530	85%	d(c|r)=0.050
	年龄40~50	20	180	90%	标准差 0.023

年龄	对某项措施看法		同意比例	Sommers d 系数
	不同意	同意		
年龄<40	360	1640	82%	d(c|r)=-0.100
年龄40~50	560	1440	72%	标准差0.013

例7

性别		病人的反应		康复比例
		未康复	康复	
男	处理组	18	2	10%
	对照组	65	15	18.75%
女	处理组	50	30	37.5%
	对照组	10	10	50%

	病人的反应		康复比例
	未康复	康复	
处理组	68	32	32%
对照组	75	25	25%

有偏比较

在对高维列联表的压缩和分层分析时，通常会出现合并的边缘表与分层的部分表会得出不同的结论，甚至出现辛普森悖论。这就要分析偏差产生的原因。

对于抽样调查数据，抽样方法的缺陷通常会产生偏差；

如例6的调查数据分析时，可以通过调查方法设计完善，使调查的男性、女性以及年轻、年老的人数差不多，从而避免混杂因素起作用，避免有偏比较。

对于实验数据，实验设计方法上的缺陷也会导致偏差。

如例7在新药疗效实验方案设计中，必须保证双盲且随机分组，使处理组和对照组中男、女病人数差不多，使“性别”不起作用，避免有偏比较。

例8：1990年人口普查死亡率（%）

年龄	文盲	小学	初中	高中、中专	大专、大学
25-29	5.499	1.112	0.724	0.396	0.315
30-34	3.354	1.319	0.897	0.497	0.272
35-39	2.055	1.392	1.038	0.680	0.380
40-44	2.940	1.893	1.511	1.123	0.657
45-49	3.198	2.924	2.490	1.721	1.433
50-54	4.537	4.518	3.952	3.210	2.367
55-59	6.835	7.310	6.289	4.926	3.879
60-64	11.686	12.467	10.505	8.031	6.507
65-69	19.710	21.228	17.975	14.344	12.200
70-	76.194	63.100	55.027	45.570	43.674
总	29.471	11.720	3.449	1.930	2.897

1990年高中、中专和大专、大学（%）

年龄	高中、中专居民人数	高中、中专人数比例%	大专、大学居民人数	大专、大学人数比例%
25-29	573426	27.41	126892	18.47
30-34	766924	36.66	88318	12.85
35-39	163206	7.80	89555	13.03
40-44	175494	8.40	97436	14.17
45-49	138857	6.64	85847	12.49
50-54	87539	4.18	75622	11.00
55-59	75524	3.61	52853	7.69
60-64	48688	2.33	31199	4.54
65-69	31860	1.52	19345	2.82
70-	30327	1.45	20218	2.94
总	2091845	100.00	687195	100.00

有偏比较

对于观察数据，有偏比较在所难免。

解决办法是找出混杂因素，按混杂因素进行分层分析，控制混杂因素的影响，以避免有偏比较。

如例8文化程度与死亡率分析中，“年龄”是混杂因素，通过按“年龄”分层，避免了有偏比较。

高维列联表的独立性

独立性的定义可以使用概率方式，也可以采用期望频数来定义。三维列联表独立性的定义与二维列联表类似。

在三维列联表中，令m_{ijk}为期望频数，若存在$\alpha_i,\beta_j,\gamma_k$使任意格(ijk)都有：

$$m_{ijk} = \alpha_i \beta_j \gamma_k$$

则称A、B、C相互独立。

其中m_{ijk}的估计为： $\hat{m}_{ijk} = \dfrac{n_{i++} n_{+j+} n_{++k}}{n^2}$

高维列联表的独立性

若存在α_i,η_{jk}使任意格都有： $m_{ijk}=\alpha_i\eta_{jk}$

则称A和(B,C)相互独立。

其中m_{ijk}的估计为：

$$\hat{m}_{ijk}=\frac{n_{i++}\eta_{+jk}}{n}$$

与A和(B,C)相互独立的情况类似，可以得到B和(A,C)以及C 和(A ,B)相互独立的期望频数的定义。

若存在ξ_{ij},ω_{ik}使任意格都有：

则称A给定后B和C条件独立。 $m_{ijk}=\xi_{ij}\omega_{ik}$

其中m_{ijk}的估计为：

$$\hat{m}_{ijk}=\frac{n_{ij+}\eta_{i+k}}{n_{i++}}$$

高维列联表的独立性

与给定A后B和C条件独立类似，可以得到给定B后A和C条件独立、给定C后A和B条件独立的期望频数定义。

根据三维列联表独立性的三种情况下不同的期望频数的估计，可以构造出似然比检验统计量：

$$G^2 = -2\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{t} n_{ijk} \ln\left(\frac{\hat{m}_{ijk}}{n_{ijk}}\right)$$

高维列联表的相关性

在三维列联表中，前面所有的独立性问题讨论完后，可以进一步分析三个变量之间的相关关系。

相关关系有两种情况：

一种是饱和模型，表示为(ABC)，即期望频数不能分解，三个属性之间不仅两两存在交互作用，而且三个之间也有交互作用；其期望频数的估计就是实际频数n_{ijk}。

另一种是齐次关联模型，表示为(AB,AC,BC)，即期望频数可分解，两两之间存在交互作用，但三个之间没有交互作用；其期望频数的估计需要使用迭代算法。

对相关关系的分析，还可以通过对数线性模型和统计软件进行分析。

高维列联表的相关性

对期望频数的迭代估计类似于二维不完备列联表中的迭代算法。

对仅有两两交互作用模型检验的原假设应为：

$$m_{ijk} = \xi_{ij}\eta_{jk}\omega_{ik}$$

采用的似然比检验统计量与独立性检验的统计量完全相同；

当检验统计量的卡方值较小时，不拒绝原假设；当卡方值较大时，拒绝原假设。

高维列联表的优比

期望频数除用来描述列联表的独立性、相关性外，还可以描述优势比。

优比不仅可以用于四格表，还可推广到一般的二维列联表。

可以取二维表的两行两列来构造一个四格表计算优比，二维表有若干个优比。

三维列联表可以按某一属性分层后形成若干二维列联表再进行优比分析。

高维列联表的优比

对于属性A,B,C相互独立时，不论按哪个属性分层，各层二维表的优比总等于1；

对于A与(B,C)相互独立时，按属性A分层后第i层二维c×t列联表的优比与i无关，故各层B与C的相合程度相同；无论按B,或C分层，这些二维列联表上的优比总等于1；

对于A给定后B和C条件独立时，按A分层的二维列联表上的优比总等于1；且按B分层各层A与C相合程度相同，按C分层各层A与B相合程度相同。

对于齐次关联模型，各层二维列联表的优比都与在第几层没有关系。

不完备高维列联表

对不完备列联表独立性的定义与完备列联表的情形类似，不同的仅仅是定义在非空格上；

以上独立性之间的关系也与完备列联表类似；

除独立性外，不完备列联表还有拟相关问题，也与完备列联表类似；

与独立性、相关性有关的检验统计量与完备列联表相类似，不同的是自由度，有的需要相应减去空格数m，有的要具体问题具体分析。