R*C 行乘列表资料分析

目录：

（1）R 行*2 列资料

i. 行为有序变量

例：冠心病患者血清血胆固醇水平（mg/100cc）

血胆固醇水平	有冠心病	无冠心病
0-199	12	307
200-219	8	246
220-259	31	439
260+	41	245

分析目的：血胆固醇水平与冠心病的发病是否有关？

分析方法：

1) 比值比分析：第一行即 0-199 组冠心病发病比值为 12/307，以此为参照，计算其它各组发病比值与第一行比值的比，以反映高胆固醇组比低胆固醇组发生冠心病的优势比。
2) 卡方检验：计算每个格子实际频数 A 与理论频数 T 差值平方与理论频数之比的累计和。反映血胆固醇水平与冠心病的发病是否有联系。
3) 相关系数：用数字表示行等级（这里为 0-199=1， 200-219=2， 220-259=3， 260+ = 4）与列等级（这里为：有冠心病=1，无冠心病=2），计算两者相关系数，以反映两者相关性强度。

输入界面：

☑ 行为有序变量　☐ 列为有序变量

	A	B	C	D	E	F
1	Cholestrol	CVD yes	CVD no			
2	0-199	12	307			
3	200-219	8	246			
4	220-259	31	439			
5	260+	41	245			
6						

开始计算

Clear　Example

输出结果：

Proportion test

Cholestrol	CVD yes	CVD no	P *	95%CI low	95%CI upp	P-value #
0-199	12	307	0.0376	0.0205	0.0665	0.0000
200-219	8	246	0.0315	0.0147	0.0634	0.0000
220-259	31	439	0.0660	0.0460	0.0933	0.0000
260+	41	245	0.1434	0.1059	0.1906	0.0000

* Event: CVD yes
# Alternative hypothesis: true P <> 0.5

比值比分析

Cholestrol	CVD yes	CVD no	OR(比值比)	95%区间下限	95%区间上限	P 值
0-199	12	307	1.0000			
200-219	8	246	0.8320	0.3348	2.0673	0.6920
220-259	31	439	1.8066	0.9133	3.5736	0.0893
260+	41	245	4.2813	2.2021	8.3237	0.0000
趋势检验*			1.7754	1.4141	2.2290	0.0000

* 阳性事件指: CVD yes
* 趋势检验自变量赋值为: 1 2 3 4
* Check .lst file for multivariable factional polynomial method.

卡方检验结果

	卡方值	自由度	P 值
	35.028451476568	3	1.20150419927845e-07
*	35.028451476568	NA	0.000499750124937531

*: p 值是用 Monte Carlo(Hope,1968)重复 2000 次计算出来的

Fisher Exact Test: P.vlaue = 6.28225393245923e-07

相关系数*

方法	相关系数	95%区间下限	95%区间上限	p 值
Spearman	-0.1448	-	-	0.0000
Pearson	-0.1403	-0.1926	-0.0872	0.0000

*行变量赋值： 1 2 3 4 ； 列变量赋值： 0 1

ii. 行为无序变量

例：三种疗法有效率

治疗方案	有效	无效
西药	57	30
中药	24	20
中西结合	130	2

分析目的：三种疗法有效率比较？

分析方法：

1）比值比分析：第一行即西组有效比值为 57/30，以此为参照，计算其它各组有效比值与西药组比值的比，以反映其它疗法比西药疗法有效性优势比。

2）卡方检验：计算每个格子实际频数 A 与理论频数 T 差值平方与理论频数之比的累计和。反映疗法不同有效率是否不同。

输入界面：

☐ 行为有序变量 ☐ 列为有序变量

	A	B	C	D	E	F
1	Treat	Yes	No			
2	Western	57	30			
3	Chinese	24	20			
4	Combine	130	2			
5						
6						

开始计算

Clear Example

输出结果：

Proportion test

Treat	Yes	No	P *	95%CI low	95%CI upp	P-value #
Western	57	30	0.6552	0.5448	0.7517	0.0053
Chinese	24	20	0.5455	0.3900	0.6931	0.6511
Combine	130	2	0.9848	0.9408	0.9974	0.0000

* Event: Yes

Alternative hypothesis: true P <> 0.5

比值比分析

Treat	Yes	No	OR(比值比)	95%区间下限	95%区间上限	P 值
Western	57	30	1.0000			
Chinese	24	20	1.5833	0.7554	3.3186	0.2235
Combine	130	2	0.0292	0.0068	0.1264	0.0000

* 阳性事件指： No

卡方检验结果

	卡方值	自由度	P 值
	57.9014309626746	2	2.6721696023283e-13
*	57.9014309626746	NA	0.000499750124937531

*: p 值是用 Monte Carlo(Hope,1968)重复 2000 次计算出来的

Fisher Exact Test: P.vlaue = 1.31220661436687e-15

（2）2 行*C 列资料

i. 列为有序资料

例：两种处理组诱发小鼠肺炎等级

处理组	小鼠肺炎等级			
	0	1	2	3
A	15	10	0	0
B	0	7	10	8

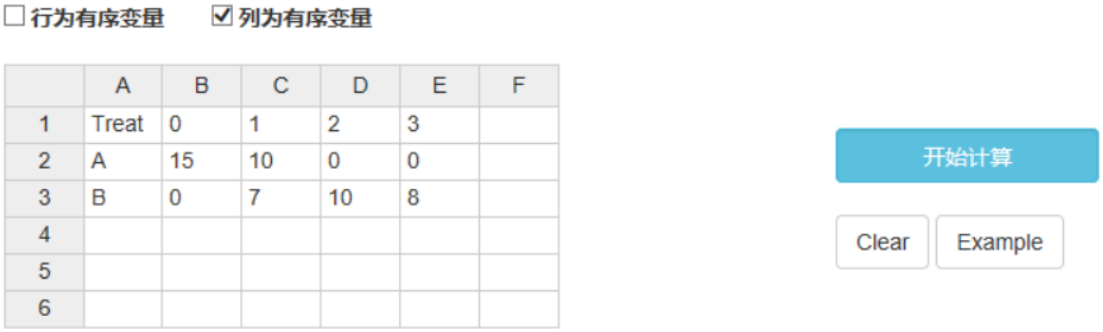
分析目的：两种处理导致的小鼠肺炎等级有无差异？

分析方法：

1） 秩和检验：结合列变量评分（此处为 0 1 2 3），比较两处理组结局变量的等级差异。

2） 卡方检验：计算每个格子实际频数 A 与理论频数 T 差值平方与理论频数之比的累计和。反映疗法不同结局是否不同，但没有结合结局变量的等级。

输入界面：



输出结果：

秩和检验

方法	统计量	自由度	p 值
Kruskal-Wallis rank sum test	17.7324131840692	1	2.54257156041189e-05

卡方检验结果

	卡方值	自由度	P 值
	33.5294117647059	3	2.49029703619515e-07
*	33.5294117647059	NA	0.000499750124937531

*: p 值是用 Monte Carlo(Hope,1968)重复 2000 次计算出来的

Fisher Exact Test: P.vlaue = 2.60020902725325e-09

ii. 列为无序资料

例：两种种治疗方案病人职业构成

治疗方案	病人职业		
	农民	工人	其它
西药	15	10	0
中药	0	7	10

分析目的：比较两种治疗方案病人职业构成是否不同？

分析方法：

1） 卡方检验：计算每个格子实际频数 A 与理论频数 T 差值平方与理论频数之比的累计和。反映两种治疗方案病人职业构成是否不同。

输入界面：

☐ 行为有序变量 ☐ 列为有序变量

	A	B	C	D	E	F
1	Treat	Farmer	Worker	Others		
2	Western	15	10	0		
3	Chinese	0	7	10		
4						
5						
6						

开始计算

Clear Example

输出结果：

卡方检验结果

	卡方值	自由度	P 值
	24.9093425605536	2	3.89946466249743e-06
*	24.9093425605536	NA	0.000499750124937531

*: p 值是用 Monte Carlo(Hope,1968)重复 2000 次计算出来的

Fisher Exact Test: P.vlaue = 2.76975049972001e-07

（3）R 行*C 列资料

i. 单向有序（行为有序或列为有序）资料

例：四种处理组诱发小鼠肺炎等级差异

	小鼠肺炎等级			
处理组	0	1	2	3
A	15	10	0	0
B	0	7	10	8
C	7	13	4	1
D	0	10	9	6

分析目的：比较各处理组间小鼠肺炎等级差异

分析方法：

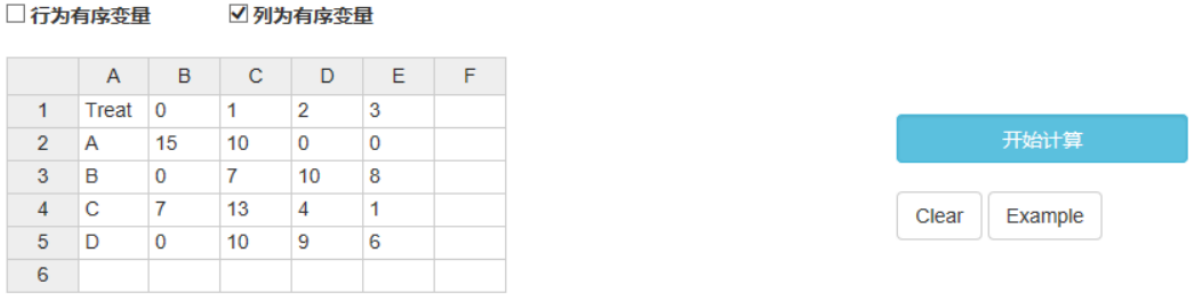
1）非参数方差分析：用来说明不同处理组小鼠肺炎等级是否有差异

2）两两组间相关系数：用数字表示列等级（这里为 0 1 2 3）与列等级（这里为 1 2，如计算 A 组与 C 组的相关性，A 组赋值为 1，C 组赋值 2），计算两者相关系数，以反映任两组相关性强度。

3）非参数方差分析后的两两比较（秩和检验）：结合列变量评分（此处为 0 1 2 3），比较任两处理组间结局变量的等级差异。

4）卡方检验：计算每个格子实际频数 A 与理论频数 T 差值平方与理论频数之比的累计和。反映处理不同，结局是否不同，未考虑结局变量的等级性。

输入界面：



输出结果：

非参数方差分析

方法	统计量	自由度	p 值
Kruskal-Wallis rank sum test	28.1271671286518	3	3.41551826268991e-06

Spearman 相关系数矩阵［相关系数（p 值）］

	A	B	C	D
A	1.0000	0.8011 (0.0000)	0.3833 (0.0060)	0.7662 (0.0000)
B	0.8011 (0.0000)	1.0000	-0.5766 (0.0000)	-0.1282 (0.3748)
C	0.3833 (0.0060)	-0.5766 (0.0000)	1.0000	0.4949 (0.0003)
D	0.7662 (0.0000)	-0.1282 (0.3748)	0.4949 (0.0003)	1.0000

Pearson 相关系数矩阵 [相关系数（p 值）]

	A	B	C	D
A	1.0000	0.7848 (0.0000)	0.3969 (0.0043)	0.7404 (0.0000)
B	0.7848 (0.0000)	1.0000	-0.5724 (0.0000)	-0.1274 (0.3781)
C	0.3969 (0.0043)	-0.5724 (0.0000)	1.0000	0.4919 (0.0003)
D	0.7404 (0.0000)	-0.1274 (0.3781)	0.4919 (0.0003)	1.0000

非参数方差分析后两两比较(multiple test after Kruskal-Wallis)

组-组	obs.dif	critical.dif	difference
1-2	36.5823754789272	22.0986009845288	1
1-3	9.23793103448276	22.5517124034683	0
1-4	31.4157088122605	22.0986009845288	1
2-3	27.3444444444444	22.9351419929878	1
2-4	5.16666666666667	22.4897570325511	0
3-4	22.1777777777778	22.9351419929878	0

critical.dif: p-value=0.05; difference: 0=差异不显著 1=差异显著
1: A; 2: B; 3: C; 4: D

卡方检验结果

	卡方值	自由度	P 值
	52.8123847167325	9	3.17233118970773e-08
*	52.8123847167325	NA	0.000499750124937531

*: p 值是用 Monte Carlo(Hope,1968)重复 2000 次计算出来的

ii. 双向有序（行和列均为有序）资料

例：两评定员对同一批切片进行等级评定结果

评定员 I 评定等级	评定员 II 评定等级			
	0	1	2	3
0	18	6	1	0
1	1	10	4	1
2	0	7	13	4
3	0	1	9	16

分析目的： 两评定员评定结果一致性如何？

分析方法：

1) 相关系数：用数字表示行等级（这里为 0 1 2 3）与列等级（这里为 0 1 2 3），计算两者相关系数，以反映相关性强度。
2) 当行数与列数相等时：

（1）关注一致性，计算 Kappa

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)},$$

其中 Pr(a) 是观察到的一致性，即（18+10+13+16）/91 ，91 是总人数。 Pr(e)是预期的一致性，首先计算对角线理论数，再合计对角线理论数/总人数。

（2）一致性图分析，计算一致性强度(strength of agreement: relation of respective area sums). 参考文献：

- Bangdiwala, S. I. (1988). The Agreement Chart. Department of Biostatistics, University of North Carolina at Chapel Hill, Institute of Statistics Mimeo Series No. 1859, http://www.stat.ncsu.edu/information/library/mimeo.archive/ISMS_1988_1859.pdf
- Bangdiwala, S. I., Ana S. Haedo, Marcela L. Natal, and Andres Villaveces. The agreement chart as an alternative to the receiver-operating characteristic curve for diagnostic tests. *Journal of Clinical Epidemiology*, 61 (9), 866-874.

（3）关注不一致性的 Mcnemar 检验（扩展至四格表之上 McNemar’s Chi-square test 又称为 Bowker 检验）：

$$\chi^2 = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}$$

就是关于对角线对称的单元格进行对比，如果 H0 成立，则两个单元格频数之差应该为 0。自由度 $=k(k-1)/2$，如果有对称的两个单元格之和为 0，则在公式中排除，自由度也相应的要减少。

输入界面：

☑ 行为有序变量　☑ 列为有序变量

	A	B	C	D	E	F
1	评定员	0	1	2	3	
2	0	18	6	1	0	
3	1	1	10	4	1	
4	2	0	7	13	4	
5	3	0	1	9	16	
6						

开始计算

Clear　Example

输出结果：

相关系数*

方法	相关系数	95%区间下限	95%区间上限	p 值
Spearman	0.6543	-	-	0.0000
Pearson	0.6547	0.5236	0.7555	0.0000

*行变量赋值： 0 1 2 3 ； 列变量赋值： 0 1 2 3

一致性分析（Kappa）

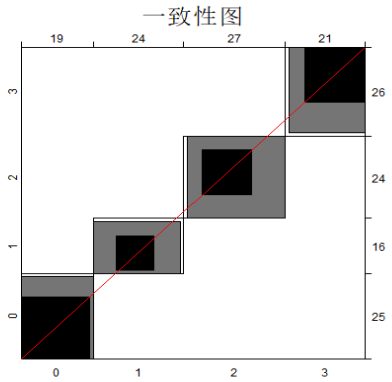
	Kappa	95%CI low	95%CI upp	p-value
Unweighted	0.5032	0.3721	0.6343	0.0000
Equal-Spacing weighted	0.6749	0.5754	0.7744	0.0000
Fleiss-Cohen weighted	0.8105	0.7345	0.8866	0.0000

The equal-spacing weights are defined by 1-|i-j|/(r-1), r:number of columns/rows, and the Fleiss-Cohen weights by 1-|i-j|^2/(r-1)^2. The latter one gives greater importance to near disagreements.

McNemar Chi-square test: Chi-square = 7.3127, df = 5, p-value = 0.198404500532792

一致性图：

Bangdiwala	Bangdiwala_Weighted	weights1	weights2
0.413541159279104	0.879417654381122	1	0.888888888888889



iii. 行和列均为无序资料

例：三种治疗方案病人职业构成

治疗方案	病人职业		
	农民	工人	其它
西药	23	10	7
中药	18	7	15
中西结合	27	9	4

分析目的： 三种治疗方案病人职业构成是否不同？

分析方法：

1） 卡方检验：计算每个格子实际频数 A 与理论频数 T 差值平方与理论频数之比的累计和。反映，治疗方案病人职业构成是否不同。

输入界面：

输出结果：

卡方检验结果

	卡方值	自由度	P 值
	9.79411764705882	4	0.0440422842259283
*	9.79411764705882	NA	0.0439780109945027

*: p 值是用 Monte Carlo(Hope,1968)重复 2000 次计算出来的

（4）1 行*C 列资料

例：某医院一周每天死亡人数是否有不同，死亡人数分布如下表：

输入界面：

☐ 行为有序变量　　☑ 列为有序变量

	A	B	C	D	E	F	G	H	I
1		1	2	3	4	5	6	7	
2	Dead	143	121	131	116	140	150	123	
3									
4									
5									
6									

输出结果：

Data

Observed	143	121	131	116	140	150	123
Expected	132	132	132	132	132	132	132

卡方检验结果

X-squared	df	P-value
7.33	6	0.291117

Spearman correlation: rho = 0.0714285714285714 P-value = 0.906349206349206

Regression (model with poisson distribution)

	RR	95%CI low	95%CI upp	P value
1	Ref.			
2	0.846	0.664	1.078	0.1762
3	0.916	0.723	1.161	0.4686
4	0.811	0.635	1.036	0.0940
5	0.979	0.776	1.236	0.8585
6	1.049	0.834	1.319	0.6826
7	0.860	0.676	1.095	0.2205
Trend (as order)	1.002	0.970	1.035	0.9083

结果解释：
设：x = 1,2,3,4,5,6,7; y = 143, 121, 131, 116, 140, 150, 123

Spearman 相关即计算 X 与 Y 的相关系数（spearman 方法）
Regression 是用的广义线性回归方程，假设 Y 的分布是 Poisson 分布，联系函数为 log.
这里有两个方程，
（1）把 X 设哑变量，即 glm(Y ~ factor(X), family=poisson(link=“log”))
（2）把 X 按连续变量处理，即 glm(Y ~ X, family=poisson(link=“log”))，也就是 Trend(as order)的结果，表示每增加一个等级，发生 dead 的发病率比增加多少。