

随机交换法计算多重比较校正的 P 值

校正多重检验 P 值的 Bonferroni 法是根据检验的次数降低显著性水平，对于 M 个独立的检验，临界显著性水平可以设定为 $0.05 / M$ 。这里的关键是 M 个独立的检验，通常我们的数据对多个应变变量或多个自变量进行的多重比较，这些应变变量之间或自变量之间相互有一定的相关性，不是完全独立的。因此用 Bonferroni 法过于保守。

随机交换 (permutation) 方法是把数据 (应变变量或自变量) 打乱，相当于重新洗牌，然后随机分配给每个观察对象，产生一个新的数据。如果原数据是 M 个自变量对某一个应变变量 (Y) 的分析，则对该应变变量 (Y) 进行随机交换；如果原数据是 M 个应变变量对某一个自变量 (X)，则对该自变量 (X) 进行随机交换。随机交换后的数据，理论上自变量与应变变量之间是没有联系的，对该数据做相应的统计检验 (M 个比较)，提取多重比较 (M 个比较) 得出的最小的 P 值，重复上述步骤 N 次 (通常 1000 次)，得出 N 个多重比较 (M 个比较) 最小的 P 值，构建最小 P 值的经验频率分布。然后把从实际数据计算出来的 P 值与该分布进行比较，以确定一个经验校正的 P 值。

参考文献：

1. Davison, A. C. & Hinkley, D. V. *Bootstrap Methods and Their Application* (Cambridge Univ. Press, 1997)
2. North, B. V., Curtis, D. & Sham, P. C. A note on calculation of empirical P values from Monte Carlo procedure. *Am. J. Hum. Genet.* 72, 498 - 499 (2003).

模块说明：

- 1) 应变变量可以是连续性或分类型的，分布类型可以是 gaussian, binomial, ordinal, poisson。应变变量还可以是依赖时间的生存状态 (0/1 变量)，分析模型用的是 cox 生存分析模型。
- 2) 该模块适用于有内部相关性的家系数据，使用 GEE 调整内部相关性。
- 3) 该模块对分类型的自变量会自动设置哑变量分析。

例：一个病例对照研究分析了 22 个 SNP 位点与疾病的关系。用随机交换法计算经验 P 值。

数据输入界面：

随机交换法计算多重比较经验P值



标题: 随机交换法计算多重比较经验P值

选择分析对象: :所有数据记录

应变量(Y)

变量名	分布	联系函数
CACO.NEW	Binomial	

自变量(X)

变量
ATIC116CS
BHMT239RQ
DNMT311IV
FOLH475HY
FOLH75YH
FTHFD793DF
FTHFD812IV
FTHFD254LP
GART752DG
GART421VI
MTHFD134RK
MTHFD653RQ
MTHFR677
MTHFR1298
MTHFS202TA
MTR919DG
MTRR595HY
MTRR22IM
MTRR350KR
MTRR415RC
MTRR175SL

Cox 模型生存分析(事件=1)

选择时间变量

或开始时间

结束时间

如用GEE

研究对象编号

内部相关类型

调整变量

变量

随机交换重抽样次数

1000

刷新

保存

查看结果

输出结果如下:

随机交换法 (permutation) 计算多重比较校正的经验 P 值

Summary table

Y	X	Coeff	Se	P.observe	P.permute	Exp(Coeff)	95% Low	95% Upp
CACO.NEW	factor(ATIC116CS)1	0.23948	0.180134	0.183698	0.999001	1.1974	0.8353	1.7163
CACO.NEW	factor(ATIC116CS)2	-0.141728	0.300902	0.637633	1	1.3511	0.3872	4.7146
CACO.NEW	factor(BHMT239RQ)1	0.179217	0.178072	0.314209	1	1.1949	0.6455	2.2121
CACO.NEW	factor(BHMT239RQ)2	0.117271	0.352517	0.739386	1	1.4226	0.334	6.06
...	...							
CACO.NEW	factor(SHMT474LF)1	-0.280023	0.187486	0.135288	0.992008	1.2062	0.9253	1.5725
CACO.NEW	factor(SHMT474LF)2	0.16656	0.285691	0.559888	1	1.3307	0.4441	3.9871

Distribution of empirical p value

0%	1%	5%	10%	25%	50%	75%	90%	95%	99%	100%
1e-04	3e-04	0.0022	0.0045	0.0101	0.0236	0.0449	0.0686	0.0849	0.1247	0.1891

注释：对应变数：CACO.NEW 进行了随机交换