

交互作用分析

一、交互作用的概念

简单地说，交互作用指当两个因素都存在时，它们的作用大于（协同）或小于（拮抗）各自作用的和。要理解交互作用首先要区别于混杂作用。

混杂作用

以吸烟（SMK）和饮酒（ALH）对收缩压（SBP）的影响为例，可以建立以下二个模型：

$$\text{模型 1: SBP} = \beta_0 + \beta_2 \text{ SMK}$$

$$\text{模型 2: SBP} = \beta_0 + \beta_1 \text{ ALH} + \beta_2 \text{ SMK}$$

假设从模型 1 估计的 SMK 的作用为 β_2' ，从模型 2 估计的 SMK 的作用为 β_2 。如吸烟与饮酒有关（假设吸烟者也多饮酒），而且饮酒与血压有关，这时可以假想两种可能：

1. 吸烟与血压无关，但因为饮酒的原因，模型 1 中的 β_2' 会显著，而模型 2 控制了 ALH 的作用后，SMK 的作用 β_2 将不显著。

2. 吸烟与血压有关，模型 1 中估计的 SMK 的作用 β_2' 一部分归功于饮酒，模型 2 估计的 β_2 是控制了 ALH 的作用后 SMK 的作用，因此 β_2' 不等于 β_2 。

是不是 β_2 不等于 β_2' 就意味着有交互作用呢？不是的，这只是意味着 β_2' 中有饮酒的混杂作用。

什么是交互作用？

$$\text{交互作用模型: SBP} = \beta_0 + \beta_1 \text{ ALH} + \beta_2 \text{ SMK} + \beta_{12} \text{ SMK} * \text{ALH}$$

根据吸烟与饮酒将研究对象分成四组，各组 SBP 的均数可用下表表示：

	不饮酒	饮酒
不吸烟	β_0	$\beta_0 + \beta_1$
吸烟	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_{12}$

吸烟与饮酒对 SBP 的影响，有无交互作用反映在 β_{12} 上，检验 β_{12} 是否等于零就是检验吸烟与饮酒对 SBP 的影响有无交互作用。而上面的模型 2 是假设 β_{12} 等于零所做的回归方程。

交互作用的评价与作用测量方法有关。以高血压发病率为例，看吸烟与饮酒对高血压发病率的影响就有两种情况。

I、相加模型：

	不饮酒	饮酒
不吸烟	I_0	$I_0 + I_a$
吸烟	$I_0 + I_s$	$I_0 + I_a + I_s + I_{sa}$

II、相乘模型：

	不饮酒	饮酒
不吸烟	I_0	$I_0 * A$
吸烟	$I_0 * S$	$I_0 * S * A * B$

相加模型检验 I_{sa} 是否等于零，相乘模型检验 B 是否等于 1，可以想象 I_{sa} 等于零时 B 不一定等于 1，因此会出现按不同的模型检验得出的结论不同。在报告交互作用检验结果时，要清楚所用的是什么模型。一般的线性回归的回归系数直接反映应变量的变化，是相加模型，而 Logistic 回归的回归系数反映比值比的变化，属相乘模型。

二、交互作用的检验

交互作用检验有两种方法，一是对交互作用项回归系数的检验（Wald test），二是比较两个回归模型，一个有交互作用项，另一个没有交互作用项，用似然比检验。本系统采用[似然比检验 \(Log likelihood ratio test\)](#) 方法。如以吸烟与饮酒两个两分类变量为例，可以形成回归方程：

$$\text{方程 1: } F(Y) = \beta_0 + \beta_1 ALH + \beta_2 SMK + \beta_{12} SMK * ALH$$

计算该方程似然数 (likelihood)，似然数表示按得出的模型抽样，获得所观察的样本的概率。它是一个很小的数，因此一般取对数表示，即 Log likelihood ，似然数可以简单地理解为拟合度。

如果我们假定吸烟与饮酒无交互作用， β_{12} 等于零，则方程为：

$$\text{方程 2: } F(Y) = \beta_0 + \beta_1 ALH + \beta_2 SMK$$

如果方程 1 和方程 2 得到的似然数没有显著差别，表明 β_{12} 是多余的，或者说 β_{12} 与零无显著性差异，吸烟与饮酒对 $f(Y)$ 无交互作用。反之，吸烟与饮酒对 $f(Y)$ 有交互作用。

三、交互作用分析

交互作用分析也可以理解为，**在分层分析基础上，对分层变量的不同层级水平上危险因素对结果变量的作用的回归系数差异进行统计学检验**。如上表中可以看出，在不吸烟组，饮酒的作用是 β_1 ，在吸烟组中饮酒的作用是 $\beta_1 + \beta_{12}$ ，如

$\beta_{12}=0$ 则表示饮酒的作用在吸烟组与不吸烟组都一样。分析交互作用主要回答的问题是：“有哪些因素影响危险因素（X）与结果变量（Y）的关系”？有没有效应修饰因子？发现效应修饰因子有助于我们进一步理解危险因素对结果变量的作用通路。

危险因素可以是连续性变量，也可以是分类型变量。本系统多要分析的可能的效应修饰因子限于分类型变量。

系统将自动检测结局变量的类型（如两分类变量、连续变量），再自动默认选择合适的回归模型（如 Logistic 回归或线性回归模型）。用户可以对分布类型和联系函数自行定义，结局变量还可以是时间依赖的生存结局。

用户可以定义表格输出格式，包括要报告的结果、行列编排、小数点位置等。

如果危险因素是分类型变量，系统将：

1. 列出危险因素与效应修饰因子的每种层级组合（联合亚组），如危险因素分 3 组，效应修饰因子分 2 组，联合亚组就有 6 组。
2. 如果结果是一个连续性的变量，统计每个联合亚组内结果变量的均数与标准差；如果结果是一个二分类的变量，统计频数(百分数)。
3. 运行两种回归模型： A 和 B
 - 模型 A 按联合亚组生成指示变量，放入模型 中（如有 6 个联合亚组，把一组作为参照组，放 入 5 个指示变量于模型中）；
 - 模型 B 不考虑危险因素与效应修饰因子的联合，分别产生指示变量放入模型中，如危险因素分 3 组，把一组作为参照，放入 2 个指示变量于模型中，效应修饰因子 分 2 组，一组为参照，放入一个指示变量 于模型中，共 3 个指示变量。

然后进行似然比检验比较模型 A 与模型 B，报告 P 值，即交互作用的 P 值。

如果危险因素是连续性变量，系统将：

1. 运行两种回归模型： A 和 B。
 - 模型 A 按效应修饰因子的每个层级产生危险 因素参数。如效应修饰因子为 SMK 分 2 组（吸烟与不吸烟），危险因素为 BMI（体重指数），产生 2 个 BMI（BMI1 与 BMI2），当 SMK=0（不吸烟）时，BMI1=BMI，BMI2=0；当 SMK=1(吸烟)时，BMI2=BMI， BMI1=0。把 BMI1 与 BMI2 同时放入模型中。
 - 模型 B 只有一个危险因素参数。

然后进行似然比检验比较模型 A 与模型 B，报告 P 值，即交互作用的 P 值。

系统将自动检测结局变量的类型（如两分类变量、连续变量），再自动默认选择合适的回归模型（如 Logistic 回归或线性回归模型）。用户可以对分布类型和联系函数自行定义。如果结局变量是时间依赖的生成结局，选“COX 模型”并给出时间变量。

例 1:

交互作用分析 ?

I. 标题:

II. 选择分析对象:

III. 选择结果变量 (Y): Cox 模型生存分析(事件=1)

变量	分布	联系函数
FEV1	Gaussian	Identity
FVC	Gaussian	Identity

IV. 选择危险因素 (X):

变量
Age, years

选择时间变量:

或, 开始时间:

结束时间:

V. 选择效应修饰因子 (分类变量):

设计行:

按分层变量:

行排序按:

再按:

再按:

选择输出内容与格式:

精确到小数点:

如用GEE

研究对象编号:

内部相关类型:

模型一: 调整变量:

模型二: 调整变量:

模型三: 调整变量:

选择每个模型分析对象:

刷新 保存 查看结果

输出结果:

交互作用检验

SMOKE ->	N	no	yes	合计	交互作用的 P 值
SEX = Male					

FEV1	361	-0.04 (-0.06, -0.03) <0.001	-0.06 (-0.06, -0.05) <0.001	-0.05 (-0.06, -0.04) <0.001	0.111
FVC	361	-0.03 (-0.04, -0.02) <0.001	-0.04 (-0.05, -0.03) <0.001	-0.04 (-0.05, -0.03) <0.001	0.085
SEX = Female					
FEV1	355	-0.03 (-0.03, -0.02) <0.001	-0.02 (-0.04, 0.01) 0.148	-0.03 (-0.03, -0.02) <0.001	0.237
FVC	355	-0.02 (-0.03, -0.02) <0.001	-0.02 (-0.04, 0.00) 0.044	-0.02 (-0.03, -0.02) <0.001	0.687
合计					
FEV1	716	-0.03 (-0.04, -0.03) <0.001	-0.05 (-0.06, -0.04) <0.001	-0.04 (-0.04, -0.03) <0.001	<0.001
FVC	716	-0.02 (-0.03, -0.02) <0.001	-0.04 (-0.05, -0.03) <0.001	-0.03 (-0.04, -0.03) <0.001	<0.001

For FEV1 FVC: 回归系数 (95%可信区间) P 值

结局变量: FEV1 和 FVC

危险因素: Age, years

分层变量: SEX

调整变量: Education, Occupation, Alcohol, Height, m 和 Weight, kg

分组合计后的分析也调整了: SEX

交互作用的 P 值由对数似然比检验比较两个相嵌模型得出

于 2015-06-02 使用《易俪统计》软件(www.empowerstats.com)和 R 软件生成。

结果解释:

以 Male、FEV1 为例，不吸烟者 AGE 每增加一岁，FEV1 下降 0.04 (95%可信区间 0.06-0.03)，吸烟者 AGE 每增加一岁，FEV1 下降 0.06，对两者差异的检验也就是吸烟是否影响 AGE 对 FEV1 的作用的交互作用检验， $p=0.111$ 。

上例是分 Male、Female 作了交互作用检验，然后做了合计的交互作用检验。每个模型都调整了 Education, Occupation, Alcohol, Height, m 和 Weight, kg。在做合计的分析时，还调整了 SEX。

例 2:

交互作用分析 ?

I. 标题:

II. 选择分析对象:

III. 选择结果变量 (Y): Cox 模型生存分析(事件=1)

变量	分布	联系函数
Systolic BP, mmhg	Gaussian	Identity
Diastolic BP, mmhg	Gaussian	Identity

IV. 选择危险因素 (X):

变量
SMOKE

选择时间变量:
 或, 开始时间:
 结束时间:

V. 选择效应修饰因子 (分类变量):

设计行:

按分层变量:
 行排序按:
 再按:
 再按:

选择输出内容与格式:
 精确到小数点:

如用GEE

模型一:
 调整变量:

 模型二:
 调整变量:
 模型三:
 调整变量:

选择每个模型分析对象:

输出结果:

交互作用检验

SMOKE	SNP1	样本量	均数+标准差	回归系数	95%可信区间	P 值	交互作用 P 值
Systolic BP, mmhg							
no	AA	311	126.30 + 20.92	0			0.354
no	AB	158	132.09 + 23.42	6.69	2.91, 10.46	<0.001	
no	BB	15	137.67 + 13.44	11.99	1.92, 22.07	0.020	
yes	AA	194	129.49 + 23.06	-6.38	-10.83, -1.93	0.005	
yes	AB	109	137.57 + 24.61	4.70	-0.34, 9.73	0.068	
yes	BB	22	140.55 + 22.01	8.04	-0.83, 16.90	0.076	
Diastolic BP, mmhg							
no	AA	311	68.53 + 10.11	0			0.290
no	AB	158	69.40 + 12.25	1.19	-0.88, 3.27	0.260	

no	BB	15	70.80 + 7.19	2.19	-3.35, 7.74	0.438	
yes	AA	194	69.22 + 10.48	-1.36	-3.81, 1.09	0.277	
yes	AB	109	72.12 + 15.03	2.46	-0.31, 5.23	0.083	
yes	BB	22	72.27 + 10.30	2.60	-2.27, 7.48	0.296	

结局变量: Systolic BP, mmhg 和 Diastolic BP, mmhg

危险因素: SMOKE 在不同的 SNP1 组内

调整变量: Age, years, SEX, Education 和 Occupation

交互作用 P 值由对数似然比检验比较两个相嵌模型得出

所有模型使用所有变量都不缺失的完整的数据

于 2015-06-02 使用《易俪统计》软件(www.empowerstats.com)和 R 软件生成。

结果解释:

以 Systolic BP (SBP) 为例, 分 SMOKE 看 SNP1 基因型间 SBP 的差异:

SMOKE=NO 的人群中, SNP1 为 AB 者比 AA 者 SBP 高 6.69mmhg, BB 者比 AA 者高 11.99mmhg;

SMOKE=YES 的人群中, SNP1 为 AB 者比 AA 者高 4.70-(-6.38)=11.08mmhg, BB 者比 AA 者高 8.04-(-6.38)=14.42mmhg。

如果 SMOKE 状况不同, SNP1 基因型间 SBP 差异不同, 这就是交互作用。但本例这种差异的显著性检验即交互作用检验不显著, $p=0.354$ 。

也可以分 SNP1 基因型看 SMOKE (yes 比 no) 对 SBP 的作用:

AA 基因型 SMOKE 的作用为 -6.38;

AB 基因型 SMOKE 的作用为 4.70-6.69=-1.99;

BB 基因型 SMOKE 的作用为 8.04-11.99=-3.95。

如果基因型不同, SMOKE 对 SBP 的作用不同, 这就交互作用, 但本例中这种交互作用的显著性检验不显著, $p=0.354$ 。