

单因素分析

单因素分析指对每个危险因素与每个结果变量一一组合进行回归分析，可以调整其它协变量。单因素分析 (univariate) 区别于单元回归方程，单元回归方程指的是方程的右边只有一个自变量。单因素分析是指当有几个危险因素的时候，对每个危险因素单独进行的分析，回归方程的右边只有这几个危险因素中的一个，但可以有其它的调整变量。单因素分析的反面就是多因素分析 (multivariate)，多因素分析是把这几个危险因素同时放入方程中，另外还可以放入其它调整变量。单因素分析的目的是对所研究的危险因素进行初步分析，同时发现潜在的混杂因素。单因素分析结果通常作为表二出现在论文中，它是分析研究危险因素对结局变量作用的第一步。论文中有时把单因素分析的结果与多因素分析的结果同时列出来。展示几个危险因素各自单独的作用及相互调整后的作用。如例 (Hao K, Chen C, Wang B, Yang J, Fang Z, Xu X. Familial aggregation of airway responsiveness: a community-based study. *Ann Epidemiol.* 2005 Nov;15(10):737-43.):

TABLE 4. Logistic GEE regression analysis of PD₂₀B* of subsequent offspring in relation to their family members

Model**	Crude		Adjusted [§]	
	OR	95%CI ^{§§}	OR	95%CI
Univariate				
Father	1.44	(0.74, 2.39)	1.62	(0.81, 2.73)
Mother	2.04	(1.08, 3.86)	2.81	(1.39, 5.70)
1st sib	2.20	(1.16, 4.13)	2.08	(1.06, 4.07)
Multivariate				
Father	1.19	(0.59, 2.38)	1.37	(0.67, 2.83)
Mother	1.75	(0.90, 3.38)	2.44	(1.19, 4.99)
1st sib	1.92	(0.98, 1.92)	1.68	(0.83, 3.40)

*PD₂₀B = 1 indicates that the individual experienced a $\geq 20\%$ drop of FEV₁ during the methacoline challenge test.

**Model: $\text{Logit}(\text{PD}_{20}\text{B}_{\text{Subsequent Offspring}}) = \beta_1 X + \beta_2 Z + \varepsilon$

Z is the matrix of environmental factors.

In univariate model:

X is [PD₂₀B_{Father} or PD₂₀B_{mother} or PD₂₀B_{1st Offspring}]

In multivariate model:

X is the matrix of [PD₂₀B_{Father}, PD₂₀B_{mother}, PD₂₀B_{1st Offspring}]

[§]Adjusted for age, sex, height, weight, and smoking status.

^{§§}Standard error is corrected by GEE model.

本模块将自动检测结局变量的类型（是两分类还是连续性变量），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）。用户也可以对分布类型和联系函数自行定义。还可以调用 Cox 模型进行生存分析或条件 logistic 回归分析。如果指定了分层变量，系统将分组统计。

回归分析的结果通常呈现回归系数或 OR，用户可以定义输出调整后的各组均数 (当结果变量是连续性变量) 或率 (当结果变量为两分类变量)，此时如果危险因素是连续性变量，本模块自动将该危险因素转换成三等分组变量。用户可以定义表格输出格式，包括要报告的结果、行列编排、小数点位数等。

例 1，使用练习数据，输入界面如下：

逐个危险因素回归分析 ?

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)

Variable	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 Male

- Age, years
- Body mass index, kg/m2
- SMOKE
- Alcohol
- Occupation
- Education

选择时间变量:

或, 开始时间:

结束时间:

V. 选择调整变量:

VI. 选择列分层变量:

VII. 选择行分层变量:

VIII. 选择输出内容与格式:

如用GEE

研究对象编号:

内部相关类型:

X. 精确到小数点:

输出结果:

单因素分析

SEX ->	Male	Female	合计
Systolic BP, mmhg			
Age, years	0.6 (0.5, 0.7) <0.001	0.9 (0.8, 1.1) <0.001	0.7 (0.7, 0.8) <0.001
Body mass index, kg/m2	-0.6 (-1.7, 0.5) 0.278	0.7 (-0.2, 1.6) 0.139	0.2 (-0.5, 0.9) 0.617
SMOKE			
no	0	0	0
yes	1.6 (-3.2, 6.4) 0.526	9.6 (1.3, 17.8) 0.024	3.7 (-0.5, 7.9) 0.086
Alcohol			
no	0	0	0
yes	1.0 (-3.8, 5.8) 0.676	13.0 (-13.2, 39.2) 0.331	1.4 (-3.4, 6.3) 0.556
Occupation			
farmer	0	0	0
others	6.6 (2.3, 10.9) 0.003	-11.7 (-16.1, -7.3) <0.001	-2.6 (-5.8, 0.5) 0.100
Education			
elementary of lower	0	0	0

middle	-3.8 (-9.9, 2.2) 0.215	-13.3 (-18.5, -8.1) <0.001	-9.1 (-13.0, -5.2) <0.001
high or above	-10.1 (-16.1, -4.2) <0.001	-11.1 (-18.3, -3.8) 0.003	-12.6 (-16.9, -8.3) <0.001
Diastolic BP, mmhg			
Age, years	0.2 (0.2, 0.3) <0.001	0.3 (0.2, 0.4) <0.001	0.3 (0.2, 0.3) <0.001
Body mass index, kg/m2	-0.4 (-1.0, 0.2) 0.173	0.3 (-0.1, 0.8) 0.145	0.0 (-0.3, 0.4) 0.838
SMOKE			
no	0	0	0
yes	1.8 (-0.7, 4.4) 0.163	4.2 (0.3, 8.2) 0.037	2.5 (0.3, 4.6) 0.023
Alcohol			
no	0	0	0
yes	1.6 (-1.0, 4.1) 0.228	6.3 (-6.1, 18.7) 0.321	1.7 (-0.7, 4.1) 0.161
Occupation			
farmer	0	0	0
others	2.5 (0.2, 4.8) 0.037	-3.5 (-5.6, -1.4) 0.001	-0.5 (-2.1, 1.0) 0.496
Education			
elementary of lower	0	0	0
middle	-0.7 (-3.9, 2.6) 0.692	-3.6 (-6.1, -1.1) 0.005	-2.0 (-4.0, 0.0) 0.047
high or above	-3.5 (-6.7, -0.3) 0.031	-0.2 (-3.7, 3.3) 0.917	-3.1 (-5.3, -0.9) 0.006

For Systolic BP, mmhg Diastolic BP, mmhg: 回归系数 (95%可信区间) P 值

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

分层变量: SEX

分组合计后的分析也调整了: SEX

于 2015-07-24 使用《易侬统计》软件 (www.empowerstats.com) 和 R 软件生成。

使用的样本量

	Male	Female	合计
Systolic BP, mmhg			
Age, years	394	399	793
Body mass index, kg/m2	394	399	793
SMOKE	392	395	787
Alcohol	389	395	784
Occupation	394	399	793
Education	393	395	788
Diastolic BP, mmhg			
Age, years	394	399	793
Body mass index, kg/m2	394	399	793
SMOKE	392	395	787
Alcohol	389	395	784
Occupation	394	399	793
Education	393	395	788

结果解释:

上列结果均是通过回归分析得出，每格内容为：回归系数（95%可信区间）P 值。

以结果变量 Systolic BP、危险因素 Age 为例，男性 (male) 结果 $0.6(0.5, 0.7) < 0.001$ 表示 Age 每增加 1 岁 (year), Systolic BP 增加 0.6 mmHg, 95% 可信区间为 0.5-0.7mmHg, $P < 0.001$ 。

以危险因素 Education 为例，男性 (male) middle 结果为： $-3.8(-9.9, 2.2) 0.215$; high or above 结果为： $-10.1(-16.1, -4.2) < 0.001$ 。分别表示：以 elementary or lower 文化程度的人为参照，文化程度为 middle 者与之相比，Systolic BP 低 3.8mmHg, 95%可信区间为-9.9 至 2.2mmHg, $P=0.215$ ；文化程度为 high or above 者与 elementary or lower 相比，Systolic BP 低 10.1mmHg, 95%可信区间为-16.1 至-4.2mmHg, $P < 0.001$ 。

例 2. 输出调整后的均数，使用练习数据 demo，输入界面如下：

单因素分析 ?

标题: 单因素分析

选择分析对象: 所有数据记录

权重

应变量(Y)

变量名	分布	联系函数
Systolic BP, mmhg	Gaussian	Identity
High BP	Binomial	Logit

暴露变量(X)

变量	等分组
Body mass index, kg/m2	.
SNP1	.
SNP2	.

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

调整变量

变量	曲线拟合
Age, years	S
SEX	.
Education	.
Occupation	.

选择列分层变量

选择行分层变量

如用GEE

研究对象编号

GEE type

精确到小数点: 0.1

输出内容与格式: adjust mean (95%CI) *

刷新 保存 查看结果

点击查看结果得输出结果如下：

	Statistics	Systolic BP, mmhg	High BP
Body mass index, kg/m2 group			
14.6667 - 20.1726	263 (33.2%)	129.2 (125.7, 132.7)	0.262 (0.195, 0.343)
20.1733 - 22.0317	265 (33.4%)	129.0 (125.6, 132.3)	0.279 (0.211, 0.358)
22.0386 - 30.8889	265 (33.4%)	132.6 (129.2, 136.0) *	0.359 (0.286, 0.441)
P trend		0.040	0.045
SNP1			
AA	508 (62.3%)	126.9 (123.9, 129.9)	0.227 (0.176, 0.286)
AB	270 (33.1%)	135.1 (131.7, 138.6) ***	0.287 (0.218, 0.367)
BB	37 (4.5%)	141.0 (134.5, 147.6) ***	0.642 (0.465, 0.787) ***
P trend		<0.001	<0.001
SNP2			
AA	428 (52.7%)	133.6 (130.5, 136.8)	0.295 (0.240, 0.358)
AB	322 (39.7%)	126.9 (123.5, 130.3) ***	0.159 (0.115, 0.216) ***
BB	62 (7.6%)	122.4 (116.9, 127.8) ***	0.099 (0.043, 0.210) **
P trend		<0.001	<0.001

表中数据: adjust Mean (95% CI)

结果变量: Systolic BP, mmhg; High BP

暴露变量: Body mass index, kg/m2 group; SNP1; SNP2

调整变量: Age, years(Smooth); SEX; Education; Occupation . Generalized additive models were applied

此表用易俪统计软件 (www.empowerstats.com) 和 R 软件生成, 生成日期: 2018-07-15

注解:

上表中, Body mass index(BMI)被分成三等分, 第一等分组 BMI 范围为 14.6667 至 20.1726, 该组 SBP 调整后的均数为 129.2, HBP 率为 26.2%。余类推。