

协变量检查与筛选

协变量检查与筛选用于确定应该调整哪些协变量以准确评价危险因素（X）对结果变量（Y）的独立作用大小。它也是用来回答：危险因素（X）对结果变量（Y）有没有独立作用，独立作用的大小是多少？

在分析 X 对 Y 的独立作用大小时，如何确定要调整哪些协变量呢？常见的错误做法是用逐步回归的方法建立回归模型确定要调整哪些协变量。之所以说这种方法是错误的，是因为逐步回归的方法是基于回归系数的 p 值确定要纳入的变量，而 p 值受到样本量的影响，这样大样本得出的结果与小样本得出的结果就可能很不同，那么小样本的结果就不可靠。正确的做法是要基于对每个协变量调整与否对 X 的回归系数的影响，来决定是否要调整该变量。

例如，要分析体重指数（BMI）对收缩压（SBP）的作用大小，固定要调整的协变量有性别（SEX）和年龄（AGE），可能需要调整的变量有吸烟（SMK）、饮酒（ALH）、文化程度（EDU）和职业（OCCU）。以 SMK 为例，系统将建立以下模型：

1. $SBP = SMK + SEX + AGE$
2. $SBP = BMI + SEX + AGE$
3. $SBP = BMI + SEX + AGE + SMK$
4. $SBP = BMI + SEX + AGE + SMK + ALH + EDU + OCCU$
5. $SBP = BMI + SEX + AGE + ALH + EDU + OCCU$

模型 1 将给出 SMK 对 SBP 的影响的 P 值，P 值小肯定需要调整，P 值大不等于不需要调整，要看下面的结果。

模型 2 是危险因素（BMI）对结果变量（SBP）基本回归模型。

模型 3 在基本模型基础上增加了 SMK，比较模型 2 和模型 3 中 BMI 的回归系数，可以计算出调整 SMK 之后 BMI 效应的改变。

模型 4 是全模型，调整了所有可能的协变量。

模型 5 在全模型基础上去掉了 SMK，比较模型 4 和模型 5 中 BMI 的回归系数，也可以计算出调整与不调整 SMK 之后 BMI 效应的改变。

通过对以上模型比较，确定是否需要调整 SMK，一般认为如果调整 SMK 导致 BMI 的回归系数改变超过 10%，则需要调整 SMK。而不能只根据 SMK 的 p 值确定是否需要调整 SMK。

系统首先对自变量做共线性筛查，共线性 $VIF = 1 / (1 - R^2)$ ，VIF 值越高共线性越高，通常取 $VIF < 10$ （剔除 $VIF \geq 10$ 的自变量）。然后运行上述 5 个模型筛选协变量。常用的筛选标准有：协变量对结果变量的回归系数 P 值 < 0.10 或引进协变量于基本模型后导致危险因素回归系数的改变超过 10%。

例（输入界面）：

协变量检查与筛选 ?

标题: 协变量检查与筛选

选择分析对象: 所有数据记录

权重:

结果变量(Y)

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

暴露变量(X) Cox 模型生存分析(事件=1)

变量: Body mass index, kg/m2

选择时间变量:

或开始时间:

结束时间:

要检查与筛选的协变量

变量	曲线拟合
SMOKE	.
SMKAMT	.
Passive smoke	.
Alcohol	.
Occupation	.
Education	.

固定要调整的变量

变量	曲线拟合
Age, years	S
SEX	.

共线性筛选(剔除VIF大于等于:)

效应修饰因子:

如用GEE

研究对象编号:

GEE type:

刷新 保存 查看结果

该输入页面定义：固定调整 AGE(曲线拟合)与 SEX，共线性筛查剔除 $VIF \geq 10$ 的协变量。

输出结果：

VIF 共线性筛查

	Step 1	Singularity check
BMI	1. 1	FALSE
AGE	1. 4	FALSE
SEX	1. 3	FALSE

SMOKE	2	TRUE
SMKAMT	1	FALSE
PSMK	1. 1	FALSE
ALH	1. 1	FALSE
OCCU. NEW	1. 1	FALSE
EDU. NEW	1. 5	FALSE

共线性筛查剔除的变量： SMOKE

Y= Systolic BP, mmhg

1. 逐个查看协变量与 Y 的关系

Covariates	N	term	beta	Se.	95%CI Low	95%CI Upp	P. value
SMKAMT	309	SMKAMT	0. 3958	0. 3197	-0. 2309	1. 0225	0. 2167
Passive smoke	788	factor (PSMK) 1	-1. 9560	1. 5704	-5. 0341	1. 1220	0. 2133
Alcohol	784	factor (ALH) 1	-0. 4750	2. 1518	-4. 6926	3. 7425	0. 8253
Occupation	793	factor (OCCU. NEW) 2	-2. 4251	1. 4651	-5. 2966	0. 4465	0. 0983
Education	788	factor (EDU. NEW) 2	-0. 3593	1. 8982	-4. 0798	3. 3612	0. 8499
		factor (EDU. NEW) 3	0. 5520	2. 2417	-3. 8417	4. 9457	0. 8056

2. 在基本模型中引进协变量与在完整模型中剔除协变量，观察 X 的回归系数的变化

X= Body mass index, kg/m²

		基本模型	完整模型	
协变量	+/- term	BMI	BMI	选出
	起始回归系数	0. 5357	0. 4340	
SMKAMT	SMKAMT	0. 4544 *	0. 5337 *	Yes
Passive smoke	factor (PSMK)	0. 4402 *	0. 5310 *	Yes
Alcohol	factor (ALH)	0. 5261	0. 4466	
Occupation	factor (OCCU. NEW)	0. 5194	0. 4672	
Education	factor (EDU. NEW)	0. 6548 *	0. 3103 *	Yes

* 表示与起始回归系数相比变化超过 10%

Y= Diastolic BP, mmhg

1. 逐个查看协变量与 Y 的关系

Covariates	N	term	beta	Se.	95%CI Low	95%CI Upp	P. value
SMKAMT	309	SMKAMT	0. 1440	0. 1709	-0. 1909	0. 4790	0. 4000
Passive smoke	788	factor (PSMK) 1	-0. 4745	0. 8595	-2. 1592	1. 2101	0. 5810
Alcohol	784	factor (ALH) 1	0. 9060	1. 1740	-1. 3952	3. 2071	0. 4405
Occupation	793	factor (OCCU. NEW) 2	-0. 7110	0. 8030	-2. 2849	0. 8629	0. 3762
Education	788	factor (EDU. NEW) 2	1. 4441	1. 0362	-0. 5868	3. 4751	0. 1638
		factor (EDU. NEW) 3	2. 3059	1. 2237	-0. 0925	4. 7043	0. 0599

2. 在基本模型中引进协变量与在完整模型中剔除协变量，观察 X 的回归系数的变化
 X= Body mass index, kg/m²

		基本模型	完整模型	
协变量	+/- term	BMI	BMI	选出
	起始回归系数	0.0258	-0.0072	
SMKAMT	SMKAMT	-0.0125 *	0.0366 *	Yes
Passive smoke	factor(PSMK)	-0.0190 *	0.0391 *	Yes
Alcohol	factor(ALH)	0.0182 *	0.0018 *	Yes
Occupation	factor(OCCU.NEW)	0.0296 *	-0.0109 *	Yes
Education	factor(EDU.NEW)	0.0825 *	-0.0675 *	Yes

* 表示与起始回归系数相比变化超过 10%

筛选出来的协变量

Y	X	选出的协变量 (标准 1)	选出的协变量 (标准 2)
Systolic BP, mmhg	Body mass index, kg/m ²	SMKAMT PSMK EDU.NEW	SMKAMT PSMK OCCU.NEW EDU.NEW
Diastolic BP, mmhg	Body mass index, kg/m ²	SMKAMT PSMK ALH OCCU.NEW EDU.NEW	SMKAMT PSMK ALH OCCU.NEW EDU.NEW

注释:

1. 标准 1: 在基本模型中引进协变量或在完整模型中剔除协变量对 X 的回归系数的影响 > 10%
2. 标准 2: 标准 1 或协变量对 Y 的回归系数 P 值 < 0.1
3. 逐个协变量对 Y 的模型、基本模型、完整模型中均调整了固定要调整的变量: Age, years SEX

结果解释:

以结果变量: Systolic BP (SBP); 危险因素: Body mass index (BMI); 协变量: Passive smoke (PSMK) 为例:

PSMK 的单因素模型 $SBP = PSMK + s(AGE) + SEX$, PSMK 的 P 值为 0.2133;

基本模型 $SBP = BMI + s(AGE) + SEX$, BMI 的回归系数为 0.5357; 在该模型中引进 PSMK 后, 即模型 $SBP = BMI + s(AGE) + SEX + PSMK$, BMI 的回归系数为 0.4402, 改变 $(0.5357 - 0.4402) / 0.5357$, 超过 10%;

完整模型 $SBP = BMI + s(AGE) + SEX + SMKAMT + PSMK + ALH + OCCU.NEW + EDU.NEW$, BMI 的回归系数为 0.4340; 在该模型中剔除 PSMK 后, 即模型 $SBP = BMI + s(AGE) + SEX + SMKAMT + ALH + OCCU.NEW + EDU.NEW$, BMI 的回归系数为 0.5310, 改变 $(0.5310 - 0.4340) / 0.4340$, 超过 10%。

右击输出文件, 打开 .lst 文件可以看到详细的各模型输出结果