

阈值效应与饱和效应分析

生物医学研究中许多研究因素在一定范围内对结果变量无影响或有正效作用，超过某一阈值后，作用大小或/和方向可发生变化，称为阈值效应。在分析研究因素（ x ）对结果变量（ y ）的作用有无阈值效应时，可先通过平滑曲线拟合观察是否有分段线性关系，然后采用分段回归模型、LRT 检验和 Bootstrap 重抽样法进行阈值效应分析。

该模块要回答的问题是：危险因素（ X ）与结果变量（ Y ）有什么样的关系？有没有阈值效应或饱和效应？参看[流行病学分析思路](#)。一般先通过平滑曲线拟合，发现危险因素与结局变量有无分段式的关系，并找到大致的转折点，然后用阈值效应与饱和效应分析模块，采用分段线性回归模型拟合数据。

转折点可以人为定义，直接“输入转折点”，也可由系统自动确定转折点。系统确定转折点的方法是采用递归实验法找出似然值最大的拟合模型。

所谓最大似然法，即从数据中寻找一个 k 值，其所得出的模型给出最大的似然值。具体操作分两步：

第一步：从 x 的 5% 百分位数开始，按 5% 逐步递增，到 95% 百分位数，共 19 个点，分别赋 k 值为这 19 个点的 x 值，找出哪个百分位数给出最大的似然值，记为 $p1$ ，并分别找出 $p1-4\%$ 与 $p1+4\%$ 百分位数所对应的 x 值，记为 $kmin$ 、 $kmax$ ，应将 k 值缩小到该范围内。

第二步：用递归方法在 $kmin$ 至 $kmax$ 之间的所有观察到的 x 取值内，找出哪个 x 作为 k 值给出最大似然值。具体方法是首先比较该范围内的 $Q1$ （25% 百分位点）、 $Q2$ （50% 百分位点）与 $Q3$ （75% 百分位点），找出哪个位点作为 k 值所给出的模型似然值最大，然后把范围缩小到该位点前后 25% 范围内，这样每次递归剔除 50% 的 x 取值。最终得出能给出最大的模型似然值的 k 值。

确定阈值置信区间（CI）：采用 Bootstrap 重抽样方法确定阈值的可信区间。即从现有数据中随机抽样，重新抽取一个同样样本量的数据 i 。抽出的个体再放回，这样在抽取的数据中，原数据中有些个体可能未被抽中，也有些个体可能被抽出多次。对抽取的数据进行阈值分析，记录所得阈值 k_i 。这样重复随机抽样 1000 次，计算出 1000 个阈值 k_i 。再计算这 1000 个 k_i 的 2.5% 与 97.5% 百分位数，即为所观察阈值的 95% CI。

系统输出表格包括一元线性回归（模型 1），分段回归模型（模型 2）。是否存在阈值效应或饱和效应，可有两种检验方法，一是检验分段斜率差是否等于零（Wald test）；二是用[似然比检验](#)比较模型 1 和模型 2。系统输出两种检验结果，通常两种检验结果一致。

如果手动输入 X 的折点，并有分层分析变量，该模块自动进行分层变量与 X 的交互作用分析，比较两段回归系数在分层变量不同层级上是否有显著差异。

例 1（输入界面）：

分段线性回归用于阈值效应或饱和效应分析 ?

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

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:

VI. 选择分层变量:

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:

选择或输入拐点(如3.8):

Bootstrap 计算拐点可信区间

V. 选择调整变量:

变量
Body mass index, kg/m2
Occupation
Education
SMOKE

刷新 保存 查看结果

输出结果：

阈值效应分析

	SEX = Male	SEX = Female	Total
模型 I			
一条线回归系数	0.7 (0.5, 0.8) <0.001	0.9 (0.7, 1.0) <0.001	0.8 (0.7, 0.9) <0.001
模型 II			
折点(K)	43.8	39.5	40.4
< K 时回归系数 1	0.0 (-0.3, 0.3) 0.996	0.3 (0.0, 0.7) 0.080	0.2 (-0.1, 0.4) 0.124
> K 时回归系数 2	1.4 (1.1, 1.8) <0.001	1.3 (1.0, 1.6) <0.001	1.3 (1.1, 1.5) <0.001
回归系数 2 与 1 的差	1.4 (0.9, 2.0) <0.001	1.0 (0.4, 1.6) 0.002	1.1 (0.7, 1.5) <0.001
对数似然比检验	0.000	0.002	0.000
折点处 Y 预测值	127.28 (123.42, 131.14)	124.18 (120.57, 127.79)	125.17 (122.54, 127.8)

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

结局变量: Systolic BP, mmhg

危险因素: Age, years

分别将每个结局变量与每个危险因素一一对应建立模型

调整变量: Body mass index, kg/m2, Occupation, Education 和 SMOKE

分层变量: SEX

分组合计后的分析也调整了: SEX

于 2015-07-25 使用《易侬统计》软件(www.empowerstats.com)和 R 软件生成。

使用的样本量

SEX	结局变量	自变量	样本量
Male	Systolic BP, mmhg	Age, years	392
Female	Systolic BP, mmhg	Age, years	395
Total	Systolic BP, mmhg	Age, years	787

结果解释:

上例中, 结果变量为 SBP(收缩压), 危险因素为 AGE(年龄), 调整 Body mass index, kg/m², Occupation, Education 和 SMOKE。首先分 SEX 分析, 然后合计分析, 合计的分析同时调整了 SEX。以合计的分析结果为例:

如果用一条直线拟合数据, 结果是 AGE 每增加一岁, SBP 增加 0.8mmHg, 95%可信区间 0.7-0.9mmHg, P<0.001;

如果用两段模型拟合数据, 找出最佳折点为 40.4 岁, AGE 在 40.4 岁前, 每增加一岁, SBP 增加 0.2mmHg, 95%可信区间-0.1 至 0.4, p=0.124; AGE 在 40.4 岁后, 每增加一岁, SBP 增加 1.3mmHg, 95%可信区间 1.1-1.5mmHg, P<0.001。两段回归系数差 1.1mmHg, 该差值的 95%可信区间为 0.7-1.5mmHg, p<0.001; 两模型比较的似然比检验 P<0.001(0.000); 找出的折点 AGE 40.4 对应的 SBP 预测值为 125.17mmHg, 95%可信区间 122.54-127.80mmHg。没有计算折点 40.4 的可信区间。

例 2 (输入界面):

分段线性回归用于阈值效应或饱和效应分析

I. 标题: 阈值效应分析

II. 选择分析对象: 所有数据记录

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

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:

VI. 选择分层变量: sex

选择输出内容与格式: β (95%CI) Pvalue / OR (95%CI) Pvalue

精确到小数点: 0.1

V. 选择调整变量:

变量: Body mass index, kg/m², Occupation, Education, SMOKE, Alcohol

Bootstrap 计算拐点可信区间

研究对象的编号:

内部相关类型:

刷新 保存 查看结果

输出结果:

阈值效应分析

	SEX = Male	SEX = Female	Total
模型 I			
一条线回归系数	0.7 (0.5, 0.8) <0.001	0.9 (0.7, 1.0) <0.001	0.8 (0.7, 0.9) <0.001
模型 II			
折点(K1,K2)	39.6, 63.6	39.6, 63.6	39.6, 63.6
< 39.6 时回归系数 1	-0.3 (-0.7, 0.1) 0.164	0.2 (-0.2, 0.6) 0.287	0.1 (-0.2, 0.3) 0.645
39.6-63.6 之间的回归系数 2	1.4 (1.0, 1.8) <0.001	1.6 (1.2, 2.0) <0.001	1.5 (1.3, 1.8) <0.001
> 63.6 时回归系数 3	-0.8 (-2.7, 1.1) 0.409	-0.3 (-2.7, 2.2) 0.837	-0.6 (-2.0, 0.9) 0.468
回归系数 1 与 2 的差	-1.7 (-2.4, -1.0) <0.001	-1.4 (-2.1, -0.7) <0.001	-1.5 (-1.9, -1.0) <0.001
回归系数 3 与 2 的差	-2.2 (-4.2, -0.3) 0.028	-1.9 (-4.3, 0.6) 0.137	-2.1 (-3.6, -0.6) 0.007
对数似然比检验	0.000	0.000	0.000

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

结局变量: Systolic BP, mmhg

危险因素: Age, years

分别将每个结局变量与每个危险因素一一对应建立模型

调整变量: Body mass index, kg/m2, Occupation, Education, SMOKE, Alcohol 和 Passive smoke

分层变量: SEX

分组合计后的分析也调整了: SEX

于 2015-07-27 使用《易侖统计》软件 (www.empowerstats.com) 和 R 软件生成。

使用的样本量

SEX	结局变量	自变量	样本量
Male	Systolic BP, mmhg	Age, years	389
Female	Systolic BP, mmhg	Age, years	395
Total	Systolic BP, mmhg	Age, years	784

结果解释:

上例中, 结果变量为 SBP(收缩压), 危险因素为 AGE (年龄), 调整 Body mass index, kg/m2, Occupation, Education 和 SMOKE。首先分 SEX 分析, 然后合计分析, 合计的分析同时调整了 SEX。以合计的分析结果为例:

如果用一条直线拟合数据, 结果是 AGE 每增加一岁, SBP 增加 0.8mmHg, 95%可信区间 0.7-0.9mmHg, $P < 0.001$;

如果用三段模型拟合数据:

AGE 在 39.6 岁前, 每增加一岁, SBP 增加 0.1mmHg, 95%可信区间-0.2 至 0.3, $p = 0.645$;

AGE 在 39.6-63.6 岁间, 每增加一岁, SBP 增加 1.5mmHg, 95%可信区间 1.3-1.8mmHg, $P < 0.001$;

AGE 在 63.6 岁后, 每增加一岁, SBP 增加-0.6mmHg, 95%可信区间-2.0 至 0.9, $p = 0.468$;

第一段与第二段回归系数差-1.5mmHg, 该差值的 95%可信区间为-1.9 至-1.0mmHg, $p < 0.001$;

第三段与第二段回归系数差-2.1mmHg, 该差值的 95%可信区间为-3.6 至-0.6mmHg, $p = 0.007$;

分三段拟合的模型与一条线拟合的模型比较的似然比检验 $P < 0.001$ (0.000)。

更多分析结果：分层变量与危险因素交互作用

"Y: SBP, X: AGE"

结果解释：首先列出结果变量 (Y) 与危险因素 (X) 变量名

Call:

```
glm(formula = SBP ~ factor(SEX) * AGE + factor(SEX) + BMI + OCCU.NEW +  
    EDU.NEW.2 + EDU.NEW.3 + SMOKE + ALH + PSMK, family = gaussian(link = "identity"),  
    data = WD, na.action = na.omit)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-51.004	-12.465	-2.208	8.927	91.614

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	96.67079	8.13833	11.878	< 2e-16 ***
factor(SEX)2	-15.30578	4.40774	-3.472	0.000544 ***
AGE	0.69018	0.07713	8.948	< 2e-16 ***
BMI	0.79029	0.31165	2.536	0.011415 *
OCCU.NEW	-3.63718	1.50218	-2.421	0.015696 *
EDU.NEW.2	1.28731	1.96556	0.655	0.512707
EDU.NEW.3	0.62242	2.31794	0.269	0.788367
SMOKE	-3.42255	1.98353	-1.725	0.084840 .
ALH	-1.70114	2.19290	-0.776	0.438137
PSMK	-0.48773	1.55724	-0.313	0.754213
factor(SEX)2:AGE	0.26934	0.10614	2.538	0.011359 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 387.0052)

Null deviance: 397779 on 783 degrees of freedom
Residual deviance: 299155 on 773 degrees of freedom
(48 observations deleted due to missingness)
AIC: 6909.2
Number of Fisher Scoring iterations: 2

结果解释：上面是用一条回归线拟合 AGE 对 SBP 的作用，分析 AGE 与分层因素 SEX 的交互作用模型。SEX 编码为 1=Male 2=Female, Male 为参照。

AGE 的回归系数 0.69018 表示 Male AGE 每增加一岁，SBP 增加 0.69018mmHg;

factor (SEX)2:AGE 项即交互作用项，回归系数 0.26934 表示 Female AGE 的回归系数与 Male AGE 的回归系数的差为 0.26934，即 Female AGE 每增加一岁，SBP 增加 0.69018+0.26934=0.95952 mmHg, $P=0.011359$ 表示男女 AGE 对 SBP 的回归系数差异有显著性，即有交互作用。

Analysis of Deviance Table

Model 1: SBP ~ factor(SEX) * AGE + factor(SEX) + BMI + OCCU.NEW + EDU.NEW.2 +

```

EDU.NEW.3 + SMOKE + ALH + PSMK
Model 2: SBP ~ AGE + factor(SEX) + BMI + OCCU.NEW + EDU.NEW.2 + EDU.NEW.3 +
SMOKE + ALH + PSMK
Resid. Df Resid. Dev Df Deviance
1      773      299155
2      774      301647 -1 -2492.0
[1] "Log likelihood ratio test: 0.00041262126677033"

```

结果解释：上面是用似然比检验，比较一条回归线拟合 AGE 对 SBP 的作用有交互作用项模型与没有交互作用项模型， $p=0.0004126$ 表示 SEX 与 AGE 对 SBP 的有显著性的交互作用。

```

Call:
glm(formula = SBP ~ factor(SEX) * tmp.X1 + factor(SEX) * tmp.X2 +
factor(SEX) * tmp.X3 + factor(SEX) * tmp.B11 + factor(SEX) +
BMI + OCCU.NEW + EDU.NEW.2 + EDU.NEW.3 + SMOKE + ALH + PSMK,
family = gaussian(link = "identity"), data = WD, na.action = na.omit)

```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-53.822 -11.264  -1.821   8.196  91.087

```

```

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      103.4556     7.9653  12.988 < 2e-16 ***
factor(SEX)2      -0.7582     3.8239  -0.198  0.84287
tmp.X1            -0.1882     0.1865  -1.009  0.31306
tmp.X2             1.3801     0.1899   7.268 8.98e-13 **
tmp.X3            -0.8095     0.9715  -0.833  0.40497
tmp.B11TRUE       39.9012     7.2688   5.489 5.49e-08 ***
BMI                1.0579     0.3078   3.437 0.00062 ***
OCCU.NEW          -0.7726     1.5523  -0.498 0.61882
EDU.NEW.2         0.8956     1.9616   0.457 0.64811
EDU.NEW.3         0.2660     2.2886   0.116 0.90751
SMOKE             -1.0027     2.0076  -0.499 0.61761
ALH               -0.1317     2.1597  -0.061 0.95140
PSMK              -2.4244     1.5806  -1.534 0.12548
factor(SEX)2:tmp.X1  0.4328     0.2532   1.709 0.08779 .
factor(SEX)2:tmp.X2  0.4222     0.2802   1.507 0.13221
factor(SEX)2:tmp.X3  0.7706     1.6000   0.482 0.63020
factor(SEX)2:tmp.B11TRUE -4.2699    10.2002  -0.419 0.67562
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

(Dispersion parameter for gaussian family taken to be 367.5264)

```

```

Null deviance: 397779 on 783 degrees of freedom
Residual deviance: 281893 on 767 degrees of freedom
(48 observations deleted due to missingness)
AIC: 6874.6

```

```

Number of Fisher Scoring iterations: 2

```

结果解释：上面是用三条回归线拟合 AGE 对 SBP 的作用，分析 AGE 与分层因素 SEX 的交互作用模型。SEX 编码为 1=Male 2=Female，Male 为参照。

tmp.X1 的回归系数-0.1882 表示 Male AGE<39.6 时，AGE 每增加一岁，SBP 减少 0.1882mmHg， $p=0.31306$ ；

tmp.X2 的回归系数 1.3801 表示 Male AGE 39.6-63.6 岁间, AGE 每增加一岁, SBP 增加 1.3801mmHg, $p=8.98 \times 10^{-13}$;

tmp.X3 的回归系数-0.8095 表示 Male AGE>63.6 时, AGE 每增加一岁, SBP 减少 0.8096mmHg, $p=0.40497$;

factor (SEX) 2:tmp.X1 项即 AGE<39.6 段交互作用项, 回归系数 0.4328 表示 Female AGE 的回归系数与 Male AGE 的回归系数的差为 0.4328, $p=0.08779$; 即 Female AGE<39.6 岁时, 每增加一岁, SBP 增加 $-0.1882+0.4328=0.2446$ mmHg;

factor (SEX) 2:tmp.X2 项即 AGE 39.6-63.6 间交互作用项, 回归系数 0.4222 表示 Female AGE 的回归系数与 Male AGE 的回归系数的差为 0.4222, $p=0.13221$; 即 Female AGE 39.6-63.6 岁间, 每增加一岁, SBP 增加 $1.3801+0.4222=1.8023$ mmHg;

factor (SEX) 2:tmp.X3 项即 AGE>63.6 段交互作用项, 回归系数 0.7706 表示 Female AGE 的回归系数与 Male AGE 的回归系数的差为 0.7706, $p=0.63020$; 即 Female AGE>63.6 岁时, 每增加一岁, SBP 增加 $-0.8096+0.7706=-0.0390$ mmHg。

Analysis of Deviance Table

```
Model 1: SBP ~ factor(SEX) * tmp.X1 + factor(SEX) * tmp.X2 + factor(SEX) *  
  tmp.X3 + factor(SEX) * tmp.B11 + factor(SEX) + BMI + OCCU.NEW +  
  EDU.NEW.2 + EDU.NEW.3 + SMOKE + ALH + PSMK
```

```
Model 2: SBP ~ tmp.X1 + tmp.X2 + tmp.X3 + tmp.B11 + factor(SEX) + BMI +  
  OCCU.NEW + EDU.NEW.2 + EDU.NEW.3 + SMOKE + ALH + PSMK
```

	Resid. Df	Resid. Dev	Df	Deviance
1	767	281893		
2	771	286414	-4	-4521

```
[1] "Log likelihood ratio test: 0.164557567749327"
```

结果解释: 上面是用似然比检验, 比较三段拟合 AGE 对 SBP 的作用有交互作用项模型与没有交互作用项模型, $p=0.16455$ 表示 SEX 与 AGE 对 SBP 的没有显著性的交互作用。