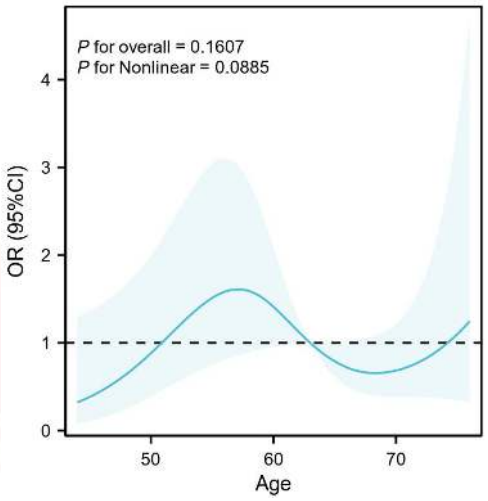
临床意义 - 诊断限制性立方样条



网址：https://www.xiantao.love

更新时间：2023.07.14

目录

基本概念 . . . 3
应用场景 . . . 3
分析流程 . . . 3
结果解读 . . . 6
数据格式 . . . 7
参数说明 . . . 10
数据处理 . . . 10
模型 . . . 11
置信区间 . . . 13
线 . . . 14
标注 . . . 15
坐标轴 . . . 17
标题文本 . . . 17
风格 . . . 18
图片 . . . 18
结果说明 . . . 19
主要结果 . . . 19
补充结果：变量情况统计表 . . . 20
补充结果：单因素 Logical 回归分析表 . . . 21
补充结果：多因素 Logical 回归分析表 . . . 22
补充结果：方差膨胀因子表 . . . 23
补充结果：模型评价 . . . 24
补充结果：非线性关联表 . . . 25
方法学 . . . 26
如何引用 . . . 27
核心代码 . . . 27
常见问题 . . . 28

基本概念

- 二分类 logistic 回归分析：当因变量是二分类变量时，常使用 logistic 回归分析自变量与因变量的关联
- 限制性立方样条图：用线来展示逻辑回归模型(Logical 模型)中比值比(OR)与变量的关系

应用场景

限制性立方样条主要用来查看/展示逻辑回归模型(Logical 模型)中比值比(OR)与变量的关系

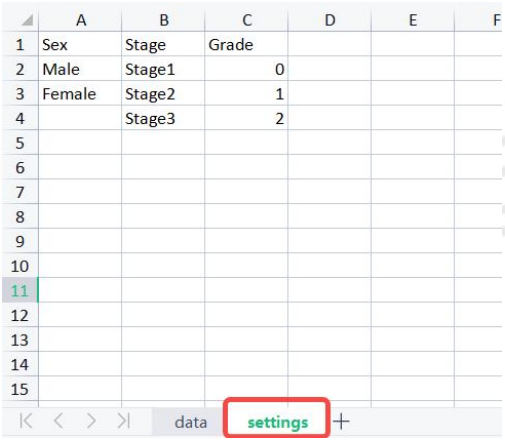
分析流程

上传数据 → 数据验证 → 数据处理(清洗) → 单因素/多因素 Logical 回归分析 → logical 回归模型评价 → 限制性立方样条分析可视化

- 数据格式：xlsx 格式
  - 第 1 列数据作为结局变量(事件发生情况)，可以是数值类型也可以是分类类型数据，需要是二分类类型，可以用（0 和 1，0 表示未发生事件，1 表示发生了事件）或（1 和 2，1 表示未发生事件，2 表示发生了事件）表示，注：第 1 列(结局变量)不能都是删失

	A	B	C	D	E	F	G
1	outcome	Age	Weight loss	Sex	Grade	Stage	Score
2	0	42	15	Male	2	Stage1	90
3	1	80		Male	0	Stage2	100
4	1	82	15	Male	0	Stage1	90
5	1	57	11	Male	0	Stage2	60
6	1	60	0	Male	2	Stage1	90
7	0	74	0	Male	2	Stage2	80
8	1	68	10	Female	0	Stage3	60
9	1	71	1	Female	2	Stage3	80
10	1	53	16	Male	1	Stage2	80
11	1	61	34	Male	0	Stage3	70
12	1	57	27	Male	1	Stage2	80
13	1	68	23	Female	1	Stage3	70
14	1	68	5	Female	0	Stage2	90

- 第 2 列以及以后为变量(支持单分类/二分类/多分类/数值(数值可以设置分组))；注意：限制性立方样条是数值类型变量在模型之间 OR 值的关系，所以意味着上传数据至少需要上传 1 列数值类型数据(除第 1 列外)，分组顺序可以在第 2 个 sheet 中设置



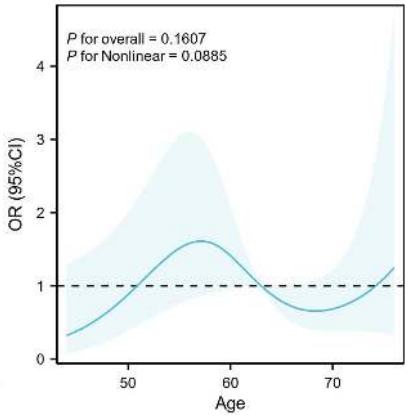
- 数据验证：
  - 先对数据第 1 列（结局）先进行基本格式要求验证（比如结局都是删失的情况、结局不是二分类情况等等）
  - 对上传数据的行数列数进行判断，需要满足分析数据的格式要求
- 数据清洗：
  - 将第 1 列（结局）不满足的数据清理掉
  - 再对其它列数据进行处理

- 如果上传数据有个 sheet 且格式贴合样本数据，可以根据第 2 个 sheet 的数据对第 1 个 sheet 的数据进行相关调整（具体可以看数据格式部分）

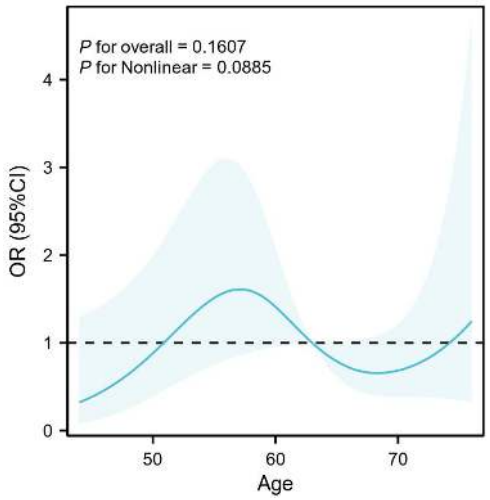
➢ 单因素/多因素 logical 回归分析：

- 构建二分类 logical 回归模型：将清洗过的数据进行 logical 模型构建
- 进行单因素/多因素 logical 回归分析

➢ 限制性立方样条分析可视化：



结果解读



- 横坐标表示变量
- 纵坐标表示模型的 Logical 回归比值比(OR 值)
- 线表示限制性立方样条曲线；阴影部分表示置信区间
- 纵坐标等于 1 位置的虚线表示，OR（比值比）等于 1 的一条参考线
- 图左上角标注表示模型非线性检验 p 值
  - P for overall：表示模型变量总的 p 值
  - P for Nonlinear：表示模型变量非线性 p 值，p < 0.05 表示选择进行分析的变量与结局(事件)之间存在非线性关系

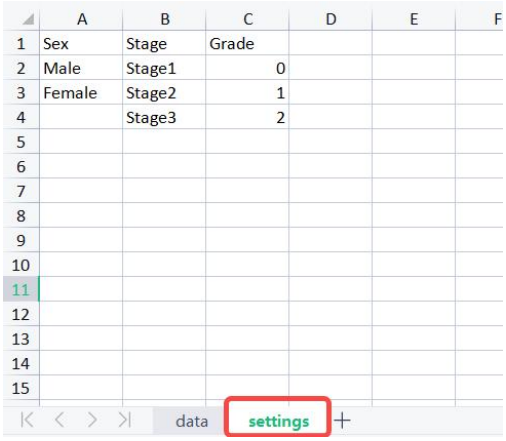
数据格式

	A	B	C	D	E	F	G
1	outcome	Age	Weight loss	Sex	Grade	Stage	Score
2	0	42	15	Male	2	Stage1	90
3	1	80		Male	0	Stage2	100
4	1	82	15	Male	0	Stage1	90
5	1	57	11	Male	0	Stage2	60
6	1	60	0	Male	2	Stage1	90
7	0	74	0	Male	2	Stage2	80
8	1	68	10	Female	0	Stage3	60
9	1	71	1	Female	2	Stage3	80
10	1	53	16	Male	1	Stage2	80
11	1	61	34	Male	0	Stage3	70
12	1	57	27	Male	1	Stage2	80
13	1	68	23	Female	1	Stage3	70
14	1	68	5	Female	0	Stage2	90

数据要求：表 1-分析数据

- 数据至少 2 列、10 行(预测变量个数*4（预测变量个数的 4 倍）行)，最多支持 41 列（40 个预测变量）和 30000 行数据
- 第 1 列表示事件发生的情况（或结局），二分类类型，可以是数值类型也可以是分类类型数据，可以用 0 和 1 表示，0 表示未发生事件，1 表示发生了事件。
- 第 2 列及以后为预测的变量，可以是数值类型，也可以是分类类型
  - 注意：限制性立方样条是数值类型变量在模型之间 OR 的关系，所以意味着上传数据至少需要上传 1 列数值类型数据（除第 1 列外）
  - 如果变量是数值变量，需以数值纳入，只要含有非数值（除空值）外，则此列有可能没有办法纳入到分析
  - 数值变量如果其分类个数 ＜10 个（如 Grade 变量只有 0 1 2）则会按照等级变量来处理

- 如果变量是等级变量，建议以具体的名字纳入，比如上图中的 Stage，也可以（类似 Grade）以数字 0 1 2 的形式纳入，但是，如果以数字编码的形式纳入，如果种类超过 5 个，需要在 excel 的表 2 中设置等级参考顺序， 否则该变量会以数值纳入（等级超过 8 个将没办法纳入）
  - 等级变量在不同等级之间的 OR 是不同的，比如结果表格中的 Stage 变量，可以看到 Stage2 和 Stage1 与 Stage4 和 Stage3 之间的 OR 是不同的。尤其要注意不要随意对一个等级资料编码为 0 1 2 3，如果在上传数据进行了此类编码，则这个变量会被认为是数值变量而产生上述数值变量的效果而出现错误。如果是进行了数字编码的等级变量，比如图中 Grade 变量，假设我们设置了 Grade 变量的等级是 0 1 2，可以在表 2 中设定该变量的等级顺序
  - 如果变量是分类变量，默认是以等级资料纳入。二分类变量以等级或者以分类资料或者数值纳入结果都是一样的。如果是多分类非等级资料，则需要以哑变量（暂不考虑）的形式纳入
  - 数值变量



	A	B	C	D	E	F
1	Sex	Stage	Grade			
2	Male	Stage1	0			
3	Female	Stage2	1			
4		Stage3	2			
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						

data settings +

数据要求:（表 2-可以不提供）：

- 对应（表 1）预测变量（分类类型）中各分类的顺序

- 比如 Stage 想要设置 Stage1, Stage2, Stage3, Stage4 的顺序，就可以如上图设置。注意，设置了等级顺序后，多因素 Cox 回归的结果都是以第一个作为参考，其他的等级顺序与第一个等级进 行对比。另外，如果在表 1 中的分类变量没有设置等级顺序，则默认以在表 1 中各个分组出现的顺序作为等级顺序。此外，如果是以 0 1 2 编码的等级变量，如果没有在这个表中进行设置，则会以数值类型纳入（可见 Grade 列）

- 如果其取值跟表 1 预测变量完全一致，则会按照其顺序对上方对应的变量分类顺序进行分析。比如 Grade 变量在表 2 中各分类的顺序为 0、1、2，与表 1 的 Grade 变量中变量名还有具体值完全一样，则会按照表 2 变量法分类的顺序进行分析，如果不是则按照表 1 中变量分类的顺序进行分析。

参数说明

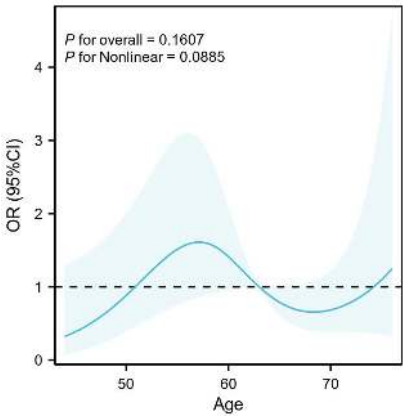
（说明：标注了颜色的为常用参数。）

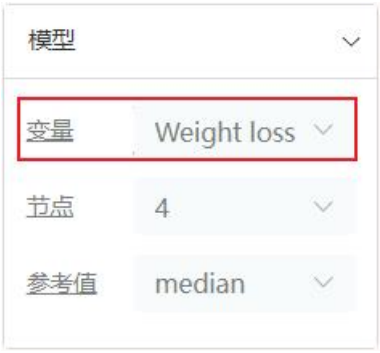
数据处理

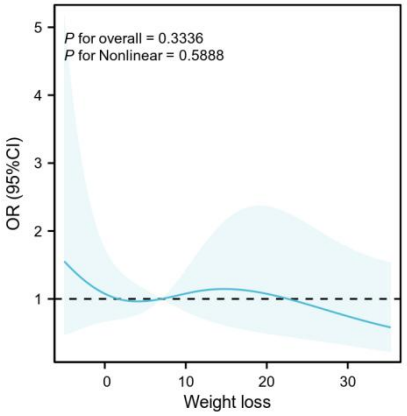
- 缺失值处理：可以选择对数据中缺失值进行处理
  - 默认为 单因素后多因素前处理变量缺失，表示在经过单因素分析之后，通过变量缺失处理在进行多因素分析
  - 还可以选择 单因素前统一处理缺失，则是在进行分析之前对全部的缺失值进行处理

模型

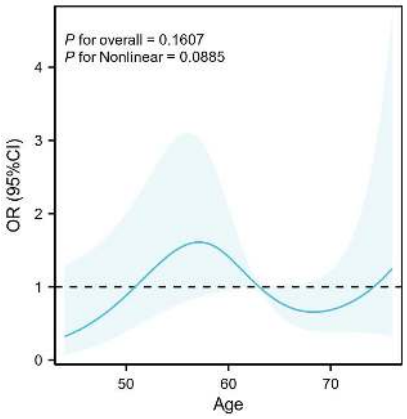
- 变量：根据上传数据特点，可以选择模型中的需要进行样条曲线可视化的单一变量(这个变量需要是数值类型的才能够进行样条曲线相关分析及可视化)，如下：

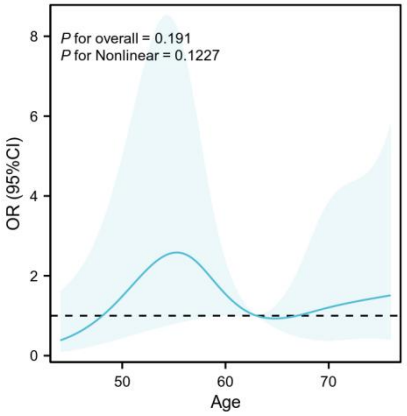


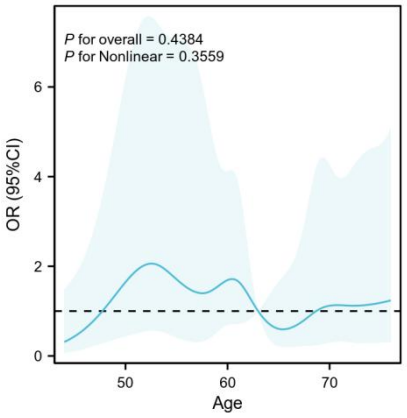




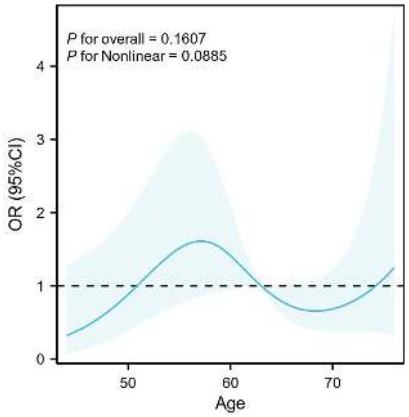
- 节点：选择模型中的需要进行样条曲线分析的节点(限制性立方样条拟合节点)；当样本量较少时可以使用 3 个节点；当样本量较大时可以使用超过 5 个节点，如下：

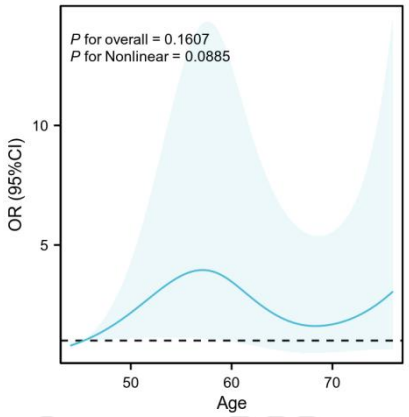






- 参考值：选择模型中的需要进行样条曲线分析的参考值，median 表示模型中的中位数，mean 表示均值，min 表示最小值，kxx 表示以模型中的拟合节点处对应值作为参考，如下：





置信区间

- 计算方法：影响补充结果中 OR 值相关置信区间计算方法。推荐选择 Wald 方法(和 SPSS 计算得到的置信区间一致)，其次是 profile 方法(MASS 包)，原本方法为传统计算方法（OR±1.96*SE），不影响可视化结果(固定使用 rms 包中 Predict 方法计算)

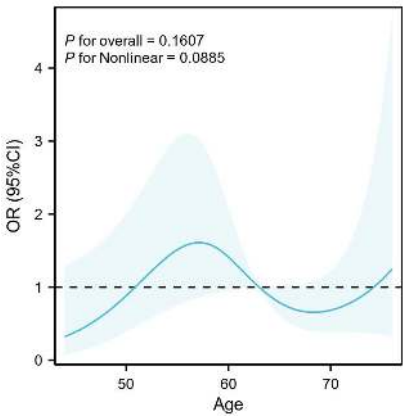
- 颜色：可以修改置信区间的颜色，置信区间默认使用传统计算方法（OR±1.96*SE）
- 不透明度：可以修改图中置信区间的不透明度，默认为 1，表示几乎不透明，1 表示完全不透明，0 表示完全透明，置信区间默认使用传统计算方法（OR±1.96*SE）

线

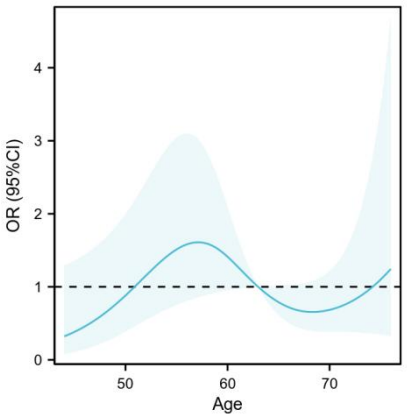
- 颜色：可以修改图中限制性立方样条曲线的颜色
- 类型：可以修改图中限制性立方样条曲线的线条类型，默认为实线，还可以选择虚线类型
- 粗细：可以修改图中限制性立方样条曲线的线条粗细，默认为 0.75pt
- 不透明度：可以修改图中限制性立方样条曲线的不透明度，默认为 1，表示几乎不透明，1 表示完全不透明，0 表示完全透明

标注

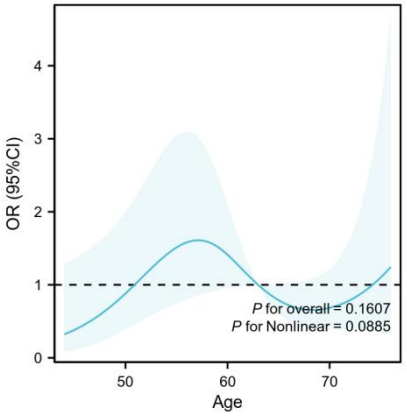
- p 值标注：可以选择是否对选择模型变量进行非线性检验 p 值标注
  - 标注：如下：
    - P for overall：表示模型变量总的 p 值
    - P for Nonlinear：表示模型变量非线性 p 值，p ＜ 0.05 表示选择进行分析的变量与死亡风险之间存在非线性关系



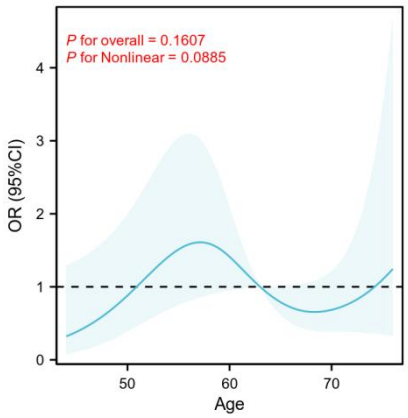
  - 不标注：如下：



- 标注位置：当选择进行 p 值标注的时候可以选择并修改 p 值标注的位置，默认为左上，还可以选择左下，右上，右下，无(等同于不展示 p 值标注)，如下：



- 标注颜色：当选择进行 p 值标注的时候可以选择并修改 p 值标注的颜色



坐标轴

- y 轴范围:可以控制 y 轴范围和刻度,可只提供 2 个值来控制范围。形如 0.1, 0.2; 0.2, 0.3(调整过大能会无作用)

标题文本

- 大标题:大标题文本
- x 轴标题:x 轴标题文本
- y 轴标题:y 轴标题文本

补充:在要换行的中间插入\n。如果需要上标,可以用两个英文输入法下的大括号括住,比如{{2}};如果需要下标,可以用两个英文输入法下的中括号括住,比如[[2]]

风格

- 外框：是否添加外框，默认添加
- 网格：是否添加网格，默认不添加
- 文字大小：控制整体文字大小，默认为 7pt

图片

- 宽度：图片横向长度，单位为 cm
- 高度：图片纵向长度，单位为 cm

字体：可以选择图片中文字的字体

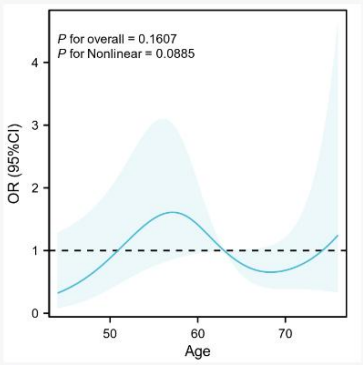
结果说明

主要结果

诊断限制性立方样条图

诊断限制性立方样条图: 用线来展示逻辑回归模型(Logical模型)中比值比(OR)与变量的关系

注意: 不同模型、同一份数据下同一个变量结果也会有所不同



诊断限制性立方样条图.pdf　诊断限制性立方样条图.tiff

· 横坐标表示变量

· 纵坐标表示模型的Logical回归比值比(OR值)

补充结果：变量情况统计表

变量情况

各个变量识别出来的类型以及是否纳入进行分析

变量	类型	分类数量	缺失数量	是否纳入分析	补充说明
outcome	分类变量	2	0	纳入	
Age	数值变量	-	0	纳入	
Weight loss	数值变量	-	14	纳入	
Sex	分类变量	2	0	纳入	
Grade	分类变量	3	0	纳入	
Stage	分类变量	3	1	纳入	
Score	数值变量	-	3	纳入	

总样本数: 228

· 如果某个分类变量的分类>10，将无法识别为分类变量/等级变量

· 如果变量的分组是以 0 1 2此类进行编码，如果分类数量<5, 会被识别为分类变量；如果>5, 会被识别为数值变量

· 如果数据中含有无穷值，无穷值会被当做缺失处理

补充说明: 单因素分析前，会先去掉 结局列中的缺失的样本(结局缺失的样本是无法纳入进行分析的)

缺失处理策略: 单因素后多因素前处理变量缺失

这里提供变量情况统计表：

- 如果某个分类变量的分类>10，将无法识别为分类变量/等级变量
- 如果变量的分组是以 0 1 2 此类进行编码，如果分类数量<5，会被识别为分类变量；如果>5，会被识别为数值变量
- 如果数据中含有无穷值，无穷值会被当做缺失处理

补充说明:

- 单因素分析前，会先去掉 结局和时间列 中的缺失的样本(时间或者结局缺失的样本是无法纳入进行分析的)
- 缺失处理策略: 单因素后多因素前处理变量缺失

补充结果：单因素 Logical 回归分析表

单因素Logistic

变量	类型	数量	OR	置信区间	p值
Age	数值变量	228	1.039	1.006 - 1.072	0.0205
Weight loss	数值变量	214	1.006	0.983 - 1.029	0.6088
Sex	等级变量	228			
Male		138	Reference		
Female		90	0.333	0.183 - 0.605	0.0003
Grade	等级变量	228			
0		40	Reference		
1		92	0.171	0.021 - 1.362	0.0953
2		96	0.024	0.003 - 0.179	0.0003
Stage	等级变量	227			
Stage1		63	Reference		
Stage2		113	1.859	0.970 - 3.560	0.0615
Stage3		51	5.270	1.961 - 14.163	0.0010

表中所有变量都会纳入到多因素中

这里提供单因素 logical 回归分析表：

- 表中所有变量都会纳入到多因素中

补充结果：多因素 Logical 回归分析表

多因素logistic

模型对应二分类结局(因变量): 1 vs. 0 (其中参考组: 0)

变量	系数β	OR	置信区间	p值
Weight loss	-0.021661	0.979	0.951 - 1.007	0.1386
Sex				
Male		Reference		
Female	-1.3773	0.252	0.116 - 0.550	0.0005
Grade				
0		Reference		
1	-1.4708	0.230	0.027 - 1.925	0.1751
2	-3.7444	0.024	0.003 - 0.191	0.0004
Stage				
Stage1		Reference		
Stage2	0.66601	1.946	0.831 - 4.559	0.1251
Stage3	1.5945	4.926	1.203 - 20.175	0.0266
Score	-0.0038113	0.996	0.967 - 1.026	0.8023

多因素logistic.xlsx

模型常数/截距(Intercept): 3.1294

原始数据一共有228个, 变量信息缺失的样本有18个, 最终纳入的样本数: 210

这里提供多因素 logical 回归分析表：

- 如果出现纳入了多因素但是对应的统计量为空的情况，说明(1)这个变量在去除变量信息缺失后某个分类数目过少(只有 1 个或者 0 个)或者是(2)存在严重共线性导致这个变量导致没办法计算
- 当如果多因素中出现 OR 异常大或者异常小时(OR 一般在 0.33-3 范围内)，说明这个变量的这个分类数量过少或者是存在共线性问题或者是数值类型变量变异比较小导致
  - (分类/等级)变量只要某个分类的 p 值满足阈值就会纳入到多因素中

补充结果：方差膨胀因子表

方差膨胀因子(VIF)

方差膨胀因子可用于分析模型中的变量是否存在多重共线性问题

变量	类型	VIF
Age	数值变量	1.1036
Weight loss	数值变量	1.1024
Sex	等级变量	
Male		Reference
Female		1.1259
Grade	等级变量	
0		Reference
1		7.4091
2		7.4974
Stage	等级变量	
Stage1		Reference
Stage2		1.3374
Stage3		1.7668

一般认为，当0 < VIF < 10，不存在多重共线性(补充: 也有认为VIF > 4就存在多重共线性); 当10 ≤ VIF < 100，存在较强的多重共线性; 当VIF >= 100或者是出现NaN，
重共线性非常严重

这里提供方差膨胀因子表：

- 一般认为，当 0 < VIF < 10，不存在多重共线性(补充：也有认为 VIF > 4 就存在多重共线性)；当 10 ≤ VIF < 100，存在较强的多重共线性；当 VIF >= 100 或者是出现 NaN，多重共线性非常严重

补充结果：模型评价

模型评价

评价方向	评价内容	统计量	p值
模型检验	似然比检验	卡方值: 78.116	1.17e-13
区分度评价	C指数	C指数: 0.859 (0.807 - 0.912)	4.58e-42
校准度评价	拟合优度检验	卡方值: 5.6004	0.6919

1. 似然比检验：若P值小于检验水准 (P < 0.05), 表示本次拟合的模型中至少有一个变量的OR值有统计学意义，即模型总体有意义
2. 模型的区分度能力采用C指数来评价，一般而言, 0.51-0.7认为是一般的准确性，0.71-0.9为中等的准确性，> 0.9为高度的准确性;
3. 校准度评价使用方法是Hosmer-Lemeshow Goodness of Fit 拟合优度检验, 一般而言, 若检验结果显示有统计学显著性(P < 0.05)，则表明模型预测值和实际观测值之间存在一定的差异，模型校准度一般; 如果P > 0.05，说明预测值与观测值没有显著差异，因此模型拟合度较好

这里提供 logical 回归模型评价表：

- 似然比检验：若 P 值小于检验水准（$P < 0.05$），表示本次拟合的模型中至少有一个变量的 OR 值有统计学意义，即模型总体有意义
- 模型的区分度能力采用 C 指数来评价，一般而言，0.51-0.7 认为是一般的准确性，0.71-0.9 为中等的准确性，> 0.9 为高度的准确性
- 校准度评价使用方法是 Hosmer-Lemeshow Goodness of Fit 拟合优度检验，一般而言，若检验结果显示有统计学显著性($P < 0.05$)，则表明模型预测值和实际观测值之间存在一定的差异，模型校准度一般；如果 $P > 0.05$，说明预测值与观测值没有显著差异，因此模型拟合度较好

补充结果：非线性关联表

非线性关联表(anova)

非线性关联表：查看变量与模型之间是否存在非线性关系
(1) 如果全局(TOTAL)满足p < 0.05，可以认为整体是有意义的(包括线性或者非线性关联)
(2) p值(Nonlinear) < 0.05为存在非线性关系

模型	统计值	自由度	p值
Age	5.1566	3	1.607e-01
Nonlinear	4.8494	2	8.850e-02
Weight loss	3.3262	1	6.820e-02
Sex	13.447	1	2.000e-04
Grade	36.899	2	9.720e-09
Stage	6.5163	2	3.850e-02
Score	0.026988	1	8.695e-01
TOTAL	45.921	10	1.480e-06

模型：模型中Age表示选择进行分析的变量
· Nonlinear：表示选择进行分析变量的非线性模型，p值(Nonlinear) < 0.05表示选择进行分析的变量与结局(事件)之间存在非线性关系
· TOTAL：整体模型，p值(TOTAL) < 0.05，可以认为整体是有意义的(包括线性或者非线性关联)
· 注意：需要关注的是，当p值(Nonlinear) < 0.05时，分析变量与结局(事件)之间存在非线性关系，说明风险增加

这里提供模型非线性关联表：

- Nonlinear：表示选择进行分析变量的非线性模型，p 值(Nonlinear) ＜ 0.05 表示选择进行分析的变量与死亡风险之间存在非线性关系
- TOTAL：整体模型，p 值(TOTAL) ＜ 0.05，可以认为整体是有意义的(包括线性或者非线性关联)
- 注意：需要关注的是，当 p 值(Nonlinear) ＜ 0.05 时，分析变量与结局(事件)之间存在非线性关系，说明风险增加

方法学

统计分析和可视化均在R 4.2.1版本中进行

涉及的R包：

1．rms[6.4.0]：用于模型拟合

2．ResourceSelection[0.3-5]:用于拟合优度检验

3．ggplot2:用于可视化

处理过程：

(1) 使用glm函数对清洗好的数据构建二分类Logistic模型

(2) 使用rms包构建限制性立方样条模型并进行相关分析

(3) 使用ggplot2包进行可视化

如何引用

生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分（Acknowledge）致谢仙桃学术（www.xiantao.love）。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。

核心代码

```r
dat <- openxlsx::read.xlsx("./二分类logistic.xlsx")
dd <- rms::datadist(dat)
options(datadist = 'dd')
# 方法一：
ddist <- rms::datadist(dat)
fit <- rms::lrm(outcome ~ rms:::rcs(Age, 4) + Weight.loss + Sex + Grade + Stage + Score, x = TRUE, y = TRUE, data = dat)
dat_plot <- rms::Predict(object = fit, name = Age, fun = exp, type = "predictions", ref.zero = TRUE, conf.int = 0.95, digits = 2)


# 方法二：
plotRCS::rcsplot(data = dat, outcome = "outcome", exposure = "Age",
                 covariates = c("Weight.loss", "Sex", "Grade", "Stage", "Score"),
                 knots = plotRCS::knot(4), ref.value = "median")
```

常见问题

1. logistic 回归分析中样本量多少才算够?

答：一般认为，每一个自变量至少要 10 例结局保证估计的可靠性。这里是结局例数，而不是整个样本例数。（如果你有 7 个自变量，那至少需要 70 例研究结 局，否则哪怕你有 1000 例，而结局的例数只有 10 例，依然显得不足。）

2. 为什么有的变量 OR 值特别大或者特别小?

答：可能有以下几个原因： 1) 如果某个分类变量中出现某个分类数量极少的情况，则可能会导致 OR 值非 常大；若是连续型自变量，可能是样本量较小或者这个自变量的变异较小（如都 在 0.1-0.2 范围）。 2) 多重共线性问题，多重共线性会产生大的标准误。 3) 空单元格，如性别与疾病的关系，所有男性都发生了疾病或都没有发生疾病， 这时候可能会出现 OR 值无穷大或为 0 的情形

3. 为什么有一些变量没有在结果上?

答：可以查看补充结果中第 1 个部分的结果，里面会说明变量纳入情况。多分类变量类别过多（>10）不会进行分析。

4. 为什么现在的 logistic 回归结果的 OR 值的置信区间和原来的不一样了?

答: 目前仙桃 logistic 相关的计算 OR 值的置信区间已经从原来的传统方法 exp(summary(model)$coefficients[2,1]+1.96*summary(model)$coefficients[2,2]) 替换成了 Wald 方法来计算置信区间, 并且提供了置信区间计算参数的方法 (当前默认会选择 Wald 方法) (Wald 方法计算得到的结果和 SPSS 中 logistic 计算得到的 置信区间是一致的), 如果计算得到旧的置信区间, 可以在置信区间参数中选择 传统方法

5. 为什么有一些变量在多因素中统计学数值为 NA?

答: 有可能是这么几种情况:

- 变量存在共线性
- 去除任一变量信息缺失后, 某个变量的某个分组变成了 0

6. 为什么文章里面的模型是放的多因素 p 值有意义的变量, 而工具却给的是多因素纳入的变量?

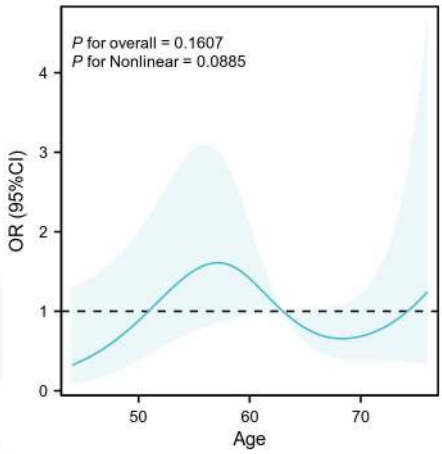
答:

首先可以明确的是, 多因素模型中的自变量就是纳入到多因素模型的所有变量, 包括进入多因素模型后 p 值没有意义的变量, 肯定不是只要多因素 p 值有意义的变量才是多因素模型的变量。

多因素模型里面正是这些所有变量在模型里面经过变量之间和混杂因素分析后才得到的每个变量的校正后的情况。

如果是提取了多因素 p 值有意义的变量再构建一个新的多因素模型，那么这些变量的系数肯定不是用的之前的那个模型，肯定是来自一个这个新的模型，而且这些在上一个多因素中 p 值有意义的变量在新的多因素模型中未必还都是 p 值有意义的。

7．可视化结果怎么看？什么时候是有意义的？



答：p 值(Nonlinear) ＜ 0.05 时，分析变量与结局(事件)之间存在非线性关系，说明死亡风险增加