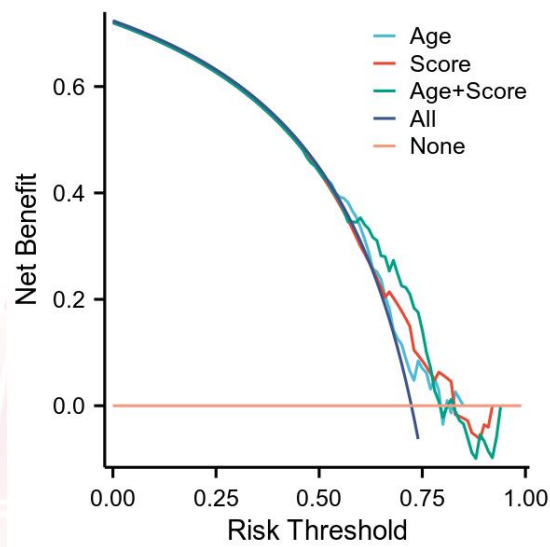


临床意义 - 诊断 DCA 图



网址: <https://www.xiantao.love>



更新时间: 2023.03.28

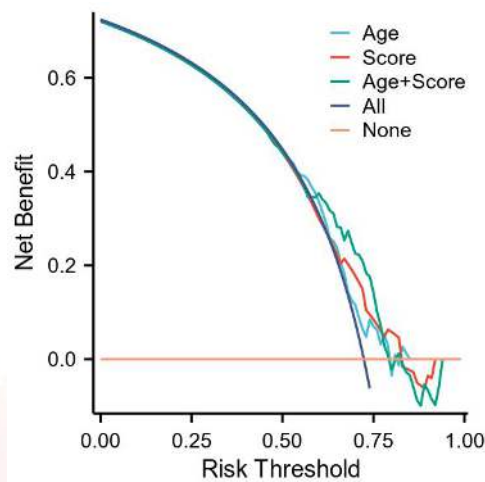
目录

基本概念	3
应用场景	3
分析流程	4
结果解读	5
数据格式	6
参数说明	7
置信区间	7
线	7
标题	8
图注	9
坐标轴	9
风格	10
图片	10
结果说明	11
主要结果	11
补充结果	12
方法学	14
如何引用	15
常见问题	16

基本概念

诊断 DCA(Decision Curve Analysis)图：用来描述随着阈概率变化，按照模型预测值对病人进行干预的情况下净收益的变化，也是一种评价诊断方法是否好用的方法。

➤ 图形构成



应用场景

可以用于评估临床预测模型、诊断试验和分子标记物

分析流程



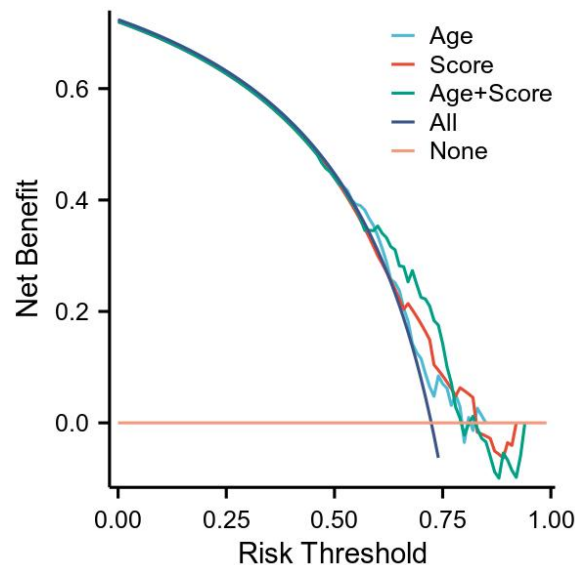
注：净收益率公式为：

真阳性率 * 患病率 - (阈值概率 / (1 - 阈值概率)) * (1 - 患病率) * 假阳性率

患病率估计方法：结局为 1 的比例



结果解读



- 横坐标为高风险概率阈值，范围是 0-1。
- 纵坐标表示净收益率，给在该高风险概率阈值下的真阳性患者施加干预的受益值和给假阳性患者施加干预的损失值之间的差值。随着高风险概率阈值的增加，根据模型结果进行干预的净收益会下降。
- 每条曲线表示每个变量（每个模型）的净收益率随着高风险概率阈值变化的情况，其中 All: 表示全部人群都干预，None: 表示全不干预，净收益一直为 0。
- All 和 None 有一个交点，表示在某个阈值水平下，对于阳性病人采取全干预和全部不干预的获益都是一样的。
- 曲线接近两条参考线（All 和 None）的模型，说明应用价值不高。Age+Score 变量在很大一个概率阈值区间内高于参考线，说明该变量模型较好。

数据格式

	A	B	C	D	E	F	G
	outcome	Age	Weight loss	Sex	Grade	Stage	Score
	0	42	15	Male		2 Stage1	90
	1	80		Male		0 Stage2	100
	1	82	15	Male		0 Stage1	90
	1	57	11	Male		0 Stage2	60
	1	60	0	Male		2 Stage1	90
	0	74	0	Male		2 Stage2	80
	1	68	10	Female		0 Stage3	60
	1	71	1	Female		2 Stage3	80
	1	53	16	Male		1 Stage2	80
	1	61	34	Male		0 Stage3	70
	1	57	27	Male		1 Stage2	80
	1	68	23	Female		1 Stage3	70
	1	68	5	Female		0 Stage2	90
	1	60	32	Male		0	70
	1	57	60	Male		0 Stage2	70
	1	67	15	Male		0 Stage2	90
	1	70	-5	Male		1 Stage2	100
	1	63	22	Male		2 Stage3	70
	1	56	10	Female		0 Stage3	60
	1	57		Male		0 Stage2	80
	1	67	17	Male		0 Stage2	80
	1	49	-8	Female		0 Stage1	70
	1	50	16	Male		0 Stage2	80
	1	58	13	Male		0 Stage1	100
	1	72	0	Male		0 Stage1	80
	1	70	6	Female		0 Stage2	80
	1	60	15	Male		0 Stage1	100

表格 1: 诊断数据 (原始数据)

	A	B	C
	Age	Score	Age+Score
	Age	Score	Age
			Score

表格 2: 变量组合

- 第一列因变量 (**必须是二分类**)，缺失值不能超过第一列长度的 85%。第 1 列中分类的前后出现的顺序会被参考的顺序，先出现的分类会被当做参考组。（影响 OR 值和置信区间计算以及图形）。
- 第 2 列及以后表示变量，可以是数值或分类变量。数据至少包含 1 个变量 (2 列)，最多不能超过 9 个变量 (10 列)，因为变量个数过多会导致曲线过于密集不利于观察。行数至少 21 行，不能超过 30000 行。
- 第 2 列以及以后可以提供变量或者是 glm 模型计算好的 linear.predictor 值来代表这个模型
- 第 2 个表可以提供想要比较的内容(1 列组合成 1 个)，不能大于 8 列，15 行

参数说明

(说明：标注了颜色的为常用参数。)

置信区间

置信区间

计算方法

Wald方法

计算方法：包含有：Wald 方法、profile 方法(MASS 包)、传统计算方法。其中，Wald 方法得到的置信区间是和 SPSS 是一致的，传统计算方法为 $(OR \pm 1.96 * SE)$ ，传统计算方法对应原本生成置信区间的方式。建议选择 Wald 方法。

线

线

颜色

样式

实线

粗细

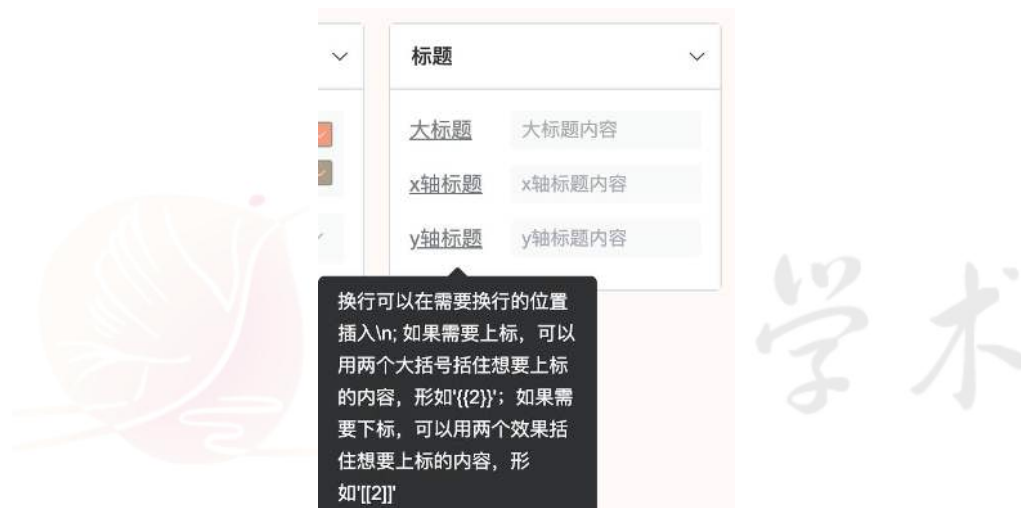
0.75pt

不透明度

1

- 颜色：拟合曲线的颜色
- 样式：两个选项，实线或者虚线，默认是实线
- 线条粗细：线条的粗细
- 不透明度：曲线的不透明度，0 为完全透明，1 为完全不透明。

标题



- 大标题：大标题文本
- X 轴标题：X 轴标题
- Y 轴标题：Y 轴标题

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]]

图注

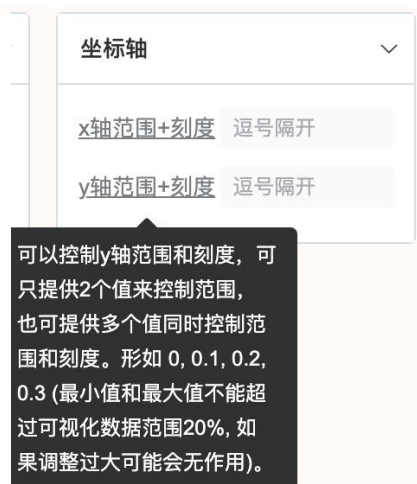


图注配置面板，包含以下选项：

- 是否展示：开关按钮，当前处于开启状态。
- 图注标题：图注标题内容
- 图注位置：默认

- 展示：是否展示图注（只有在多个组时才会有）
- 图注标题：给图注添加标题
- 图注位置：可选上、下，右下等，默认为右上

坐标轴



坐标轴配置面板，包含以下选项：

- x轴范围+刻度 逗号隔开
- y轴范围+刻度 逗号隔开

可以控制y轴范围和刻度，可只提供2个值来控制范围，也可提供多个值同时控制范围和刻度。形如 0, 0.1, 0.2, 0.3 (最小值和最大值不能超过可视化数据范围20%, 如果调整过大可能会无作用)。

- x 轴范围+刻度：控制 x 轴的刻度范围
- y 轴范围+刻度：控制 y 轴的刻度范围

风格



- 边框：是否显示主图边框
- 网格：是否添加网格
- 文字大小：图中的文字部分的大小（包括标签文字和刻度数），默认是 7pt

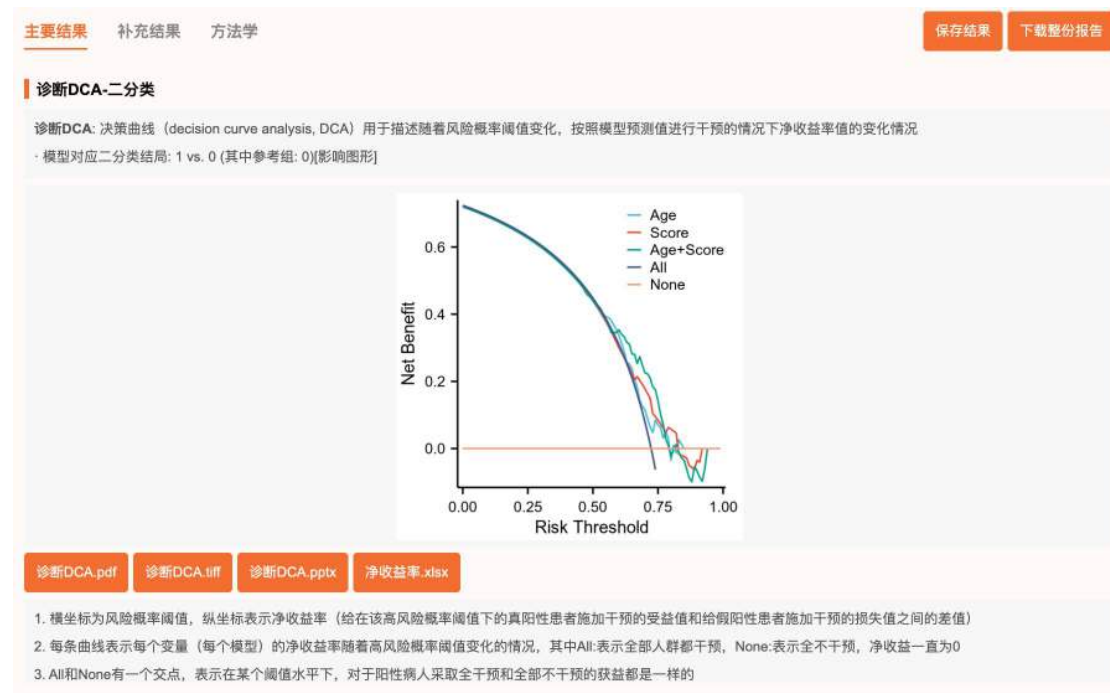
图片



- 宽度：图片横向长度，单位为 cm
- 高度：图片纵向长度，单位为 cm

结果说明

主要结果



主要结果格式为图片格式, 提供 PDF、TIFF、PPTX 格式下载。

补充结果

变量情况表：输入数据的变量基本信息及是否纳入分析说明。

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

logistic 分析的结果, OR 值, 95%置信区间和 P 值

Logistic模型: Age						
模型对应二分类结局(因变量): 1 vs. 0 (其中参考组: 0)						
变量	类型	数量	系数 β	OR	置信区间	p值
Age	数值变量	228	0.03786	1.039	1.006 - 1.072	0.0205

模型常数/截距(Intercept): -1.3783

原始数据一共有228个, 变量信息缺失的样本有0个, 最终纳入的样本数: 228

备注: 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。

备注: 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

△ 模型全局情况:

·· AIC值: 267.310 (一般来说AIC越小, 模型拟合越好)

Logistic模型: Score						
模型对应二分类结局(因变量): 1 vs. 0 (其中参考组: 0)						
变量	类型	数量	系数 β	OR	置信区间	p值
Score	数值变量	225	-0.02848	0.972	0.951 - 0.994	0.0112

模型常数/截距(Intercept): 3.2598

原始数据一共有228个, 变量信息缺失的样本有3个, 最终纳入的样本数: 225

备注: 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。

备注: 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

△ 模型全局情况:

·· AIC值: 263.866 (一般来说AIC越小, 模型拟合越好)

Logistic模型: Age+Score

模型对应二分类结局(因变量): 1 vs. 0 (其中参考组: 0)

变量	类型	数量	系数 β	OR	置信区间	p值
Age	数值变量	225	0.033362	1.034	1.001 - 1.068	0.0444
Score	数值变量	225	-0.02699	0.973	0.952 - 0.995	0.0186

模型常数/截距(Intercept): 1.0788

原始数据一共有228个, 变量信息缺失的样本有3个, 最终纳入的样本数: 225

备注: 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。

备注: 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

△ 模型全局情况:

·· AIC值: 261.765 (一般来说AIC越小, 模型拟合越好)



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：

- rmda 净收益率的计算
- ggplot2 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。



常见问题

1. 诊断 DCA 图适用于所有类型数据吗?

答：严格来说诊断 DCA 模块适应的是队列研究数据，对于病例对照数据，它涉及分层抽样，还需要考虑人群患病率，计算 NB 的方法会不太一样。

2. 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 $\exp(\text{summary(model)}\$coefficients[2,1]+1.96*\text{summary(model)}\$coefficients[2,2])$ 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。

3. 如何更改结局二分类中的顺序（参考组）?

主要结果

补充结果

方法学

诊断DCA-二分类

诊断DCA: 决策曲线 (decision curve analysis, DCA) 用于描述随着风险模型对应二分类结局: 1 vs. 0 (其中参考组: 0)[影响图形]

答：

结局变量二分类的顺序按照数据中第 1 列中分类的出现顺序来，比如下面的数据先出现的是 0，然后是 1，所以参考组就是取的第 1 个出现的组。

	A	B	C	D	E	F	G	H
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	

如果要调整这个顺序，可以自己在上传数据里面把行换一下，然后再上传数据即可：

	A	B	C	D	E	F	G	H
1	outcome	Age	Weight loss	Sex	Grade	Stage	Score	
2	1	80		Male	0	Stage2	100	
3	0	42	15	Male	2	Stage1	90	
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	

