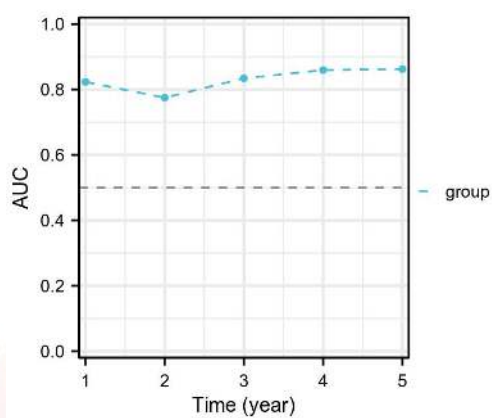


临床意义 - 时间依赖 AUC



网址: <https://www.xiantao.love>



更新时间: 2023.02.13

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	8
参数说明	10
时间点	10
置信区间	12
线	13
点	14
标题文本	15
图注	16
坐标轴	16
风格	17
图片	18
结果说明	19
主要结果：单个预测变量多个预测时间点	19
主要结果：多个预测变量多个预测时间点	19
补充结果：单个预测变量多个预测时间点的情况	20
补充结果：多个预测变量多个预测时间点的情况（且计算置信区间的前提下）	22
方法学	24
如何引用	25
常见问题	26

基本概念

- 受试者工作特征曲线 (receiver operating characteristic curve, ROC) : 一种坐标图式的分析工具, 可用于选择最佳的模型、舍弃次佳的模型或者在同一模型中设定最佳阈值。比如, 一些新标志物的预测模型会用 ROC 曲线来评价其性能。
- 时间依赖 ROC: ROC 曲线在生存资料类型的模型中的一种特殊的应用, 对于给定的时间点, 如果一个样本在时间 0 和时间 t 之间发病, 那么其会被划分到病例组, 否则则划分到对照组。对于每一个自变量不同的阈值划分分组最终就能得到每个阈值下的敏感度、特异度, 获得最佳阈值并且评估生存资料类型模型的好坏。
- 时间依赖 AUC: 用折线的形式来展示不同变量在不同时间下其对应 ROC 曲线下面积 (AUC) 的变化情况

应用场景

时间依赖 AUC 曲线主要用于连续变量来预测带有生存资料的结局事件, 找到连续变量最佳的 cut-off 值; 或者用于评估变量对于带有生存资料的结局的预测效能; 或英语评估预测变量之间在不同预测时间下的情况

分析流程

上传数据 ➡ 数据处理(清洗) ➡ AUC 相关分析 ➡ AUC 曲线可视化

- 数据格式: xlsx / text / csv

- 第 1 列数据作为结局变量(事件发生情况)，需要是数值类型数据，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
- 第 2 列数据作为时间变量(具体时间/生存时间，必须以天作为单位，并且时间要长于 1 年以上)，需要是数值类型数据，注：第 2 列(时间变量)不能都是同一个时间，并且不能出现小于 0(负数)和非数值的情况

	A	B	C
1	event	time	group
2	1	400	14.5
3	0	4500	1.1
4	1	1012	1.4
5	1	1925	1.8
6	0	1504	3.4
7	1	2503	0.8
8	0	1832	1
9	1	2466	0.3
10	1	2400	3.2
11	1	51	12.6
12	1	3762	1.4
13	1	304	3.6
14	0	3577	0.7
15	1	1217	0.8
16	1	3584	0.8

- 第 3 列开始直至后面每 1 列都代表一个变量/样本/.....，必须是数值类型数据
- 注：
 - ◆ 数据都是要是数值类型的数据
 - ◆ 不管是第 1 列（事件）、第 2 列（时间）还是其他列变量，都不能是一个值(单分类)，也不能上传非法数值(非法字符等)
 - ◆
- 数据处理：分别对第 1 列（事件）、第 2 列（时间）、第 3 列开始后的所有变量进行清洗（去除掉数据中的非数值或者不符合条件的数据）

➤ AUC 相关分析：

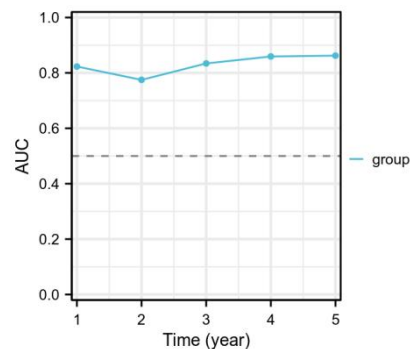
- 统计描述：在预测的各个不同时间点，统计各变量的生存情况
- ROC 结果：在预测的各个不同时间点，计算各变量的最佳阈值（cut-off 值）、敏感度、特异度等
- AUC 信息：在预测的各个不同时间点，计算各变量 ROC 曲线下面积（AUC）、累计生存率、累计发病率等
-

➤ AUC 可视化

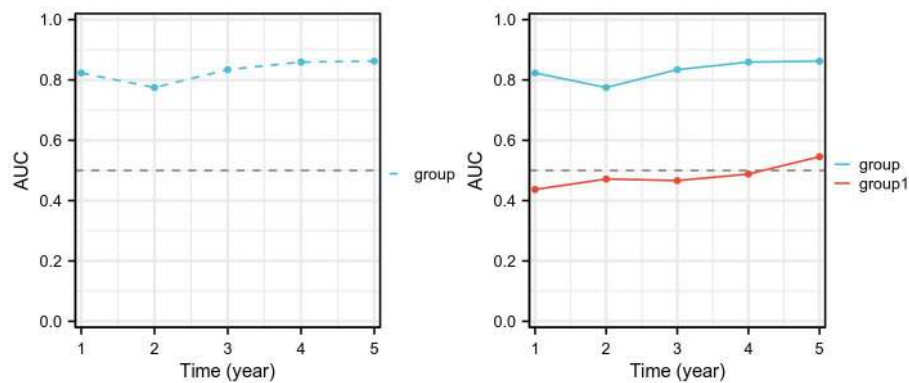
- 将分析所得到变量（group）的各个不同时间点对应的 AUC（曲线下面积）值进行可视化，结果如下：

- ◆ 模块（样本数据）默认选择（1,2,3,4,5（年））来作为预后时间，其各时间点对应的 AUC 值为：

变量	时间(年)	天数	曲线下面积(AUC)	累计生存率	累计发病率
group	1	365	0.82298	0.92823	0.07177
group	2	730	0.77483	0.88022	0.11978
group	3	1095	0.83417	0.80133	0.1961
group	4	1460	0.85934	0.75158	0.24842
group	5	1825	0.86228	0.70287	0.29713



结果解读



- 横坐标表示预测时间
- 纵坐标表示曲线下面积
- 每一条折线表示：每一个预测变量在参数设定的时间点内的 AUC 值的变化情况
 - 一条折线即表示一个预测变量
 - 如上结果所示：左侧为一个预测变量在 5 个预测时间点的结果，右侧为两个预测变量在 5 个预测时间点的不同结果
- 每一个点表示：每一个预测变量在参数设定的时间点内各时间节点的 AUC 取值
 - 每一个点表示预测变量在某个预测时间对应的 AUC 值

补充：

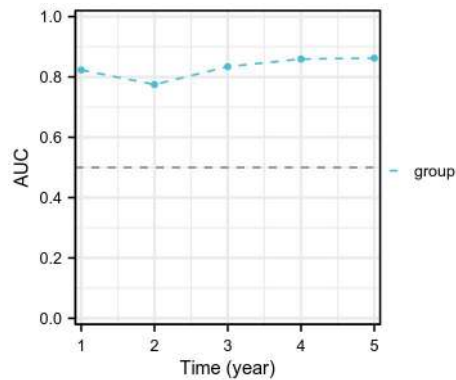
- 当一个变量(危险因素)的值是促进事件发生的趋势时,该分子的 AUC 会 >0.5 , 此时面积越大(AUC 值越接近于 1)说明预后效能越好
 - AUC 在 $0.5 \sim 0.7$ 时，预后效能较低
 - AUC 在 $0.7 \sim 0.9$ 时，有一定预后效能

- AUC 在 0.9 以上时，预后效能较好
- 当一个变量(保护因素)的值与事件发生的趋势相反时，则这个分子的 AUC 会 < 0.5 ，此时面积越小(AUC 值越接近于 0)说明预后效能越好
- 如果某个时间点对应的曲线下面积为 NA，表示该时间点下曲线下面积不存在；如果某个时间点的事件数为 0，则该时间点无法进行相关的 ROC 分析和可视化

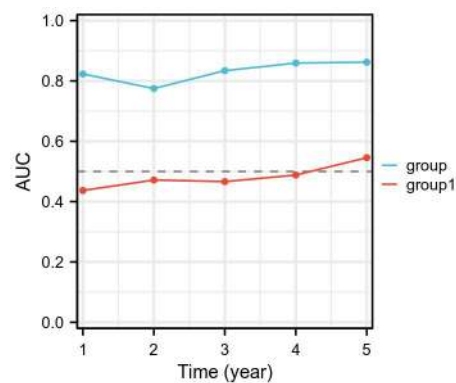


数据格式

	A	B	C
1	event	time	group
2	1	400	14.5
3	0	4500	1.1
4	1	1012	1.4
5	1	1925	1.8
6	0	1504	3.4
7	1	2503	0.8
8	0	1832	1
9	1	2466	0.3
10	1	2400	3.2
11	1	51	12.6
12	1	3762	1.4
13	1	304	3.6
14	0	3577	0.7
15	1	1217	0.8



	A	B	C	D
1	event	time	group	group1
2	1	400	14.5	1
3	0	4500	1.1	2
4	1	1012	1.4	3
5	1	1925	1.8	4
6	0	1504	3.4	5
7	1	2503	0.8	6
8	0	1832	1	7
9	1	2466	0.3	8
10	1	2400	3.2	9
11	1	51	12.6	10
12	1	3762	1.4	11
13	1	304	3.6	12
14	0	3577	0.7	13
15	1	1217	0.8	14



数据要求:

- 数据至少 3 列 30 行
- 最多支持 8 列和 5000 行数据
- 第一列是事件发生情况，用 0 和 1 表示，0 表示未发生事件，1 表示发生了事件。例如，事件可以定义为死亡，当受试发生了死亡，该受试的事件就定义为 1，当受试未发生死亡（删失），该受试的事件就定义为 0
- 第二列是具体时间，必须以天作为单位，并且时间要长于 1 年以上
- 第三列及以后为预测的变量

- 第一组结果为单个预测变量与多个预测时间点的分析结果，第二组为多个预测变量与多个预测时间点的分析结果



参数说明

(说明：标注了颜色的为常用参数。)

时间点

- 时间点：单位为年的数值，用逗号隔开，可以使用英文下的逗号也可使用中文下的逗号

- 至少需要输入 3 个不同时间节点作为预测时间；如下：

- ◆ 输入 2 个不同的预测时间

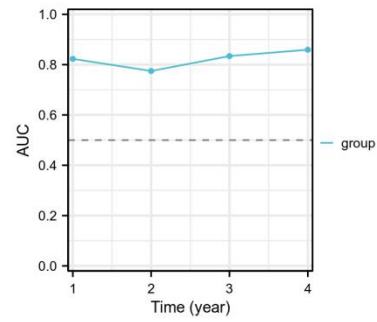
- ◆ 输入 2 个重复时间

- ◆ 输入多(≥ 3)个重复时间

预测时间

时间点 1,2,1,3, 1,4

时间单位 年

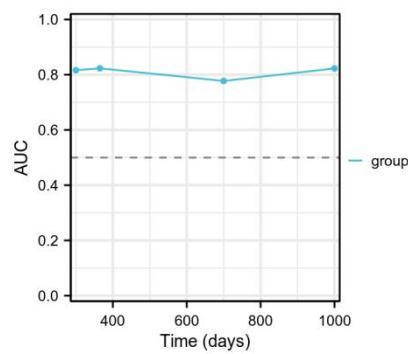


- **时间单位**：可以选择上传数据预测时间列的单位，默认以年为单位，可以选择月、天为单位，如下：

预测时间

时间点 300,365,700,1000

时间单位 日



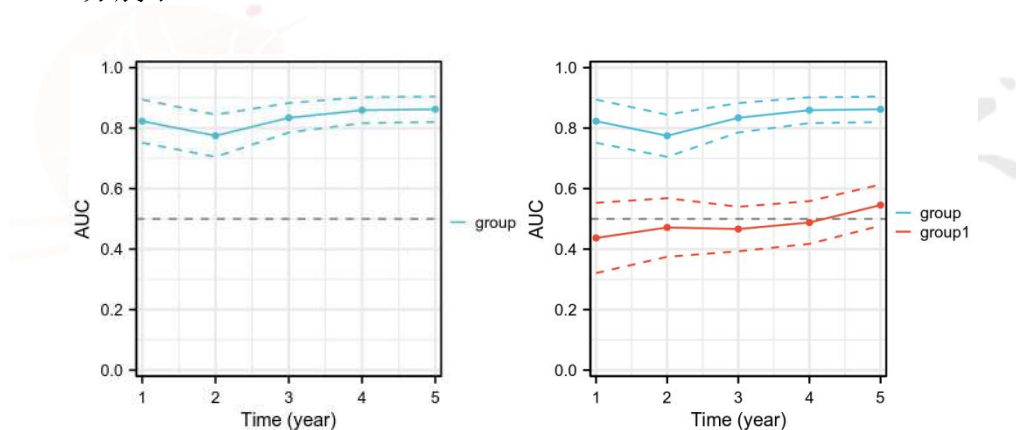
置信区间

置信区间

是否计算

- 是否计算：可以选择是否需要进行 AUC（曲线下面积）置信区间相关计算、或在多个预测变量进行分析时可以计算其组间检验，对于数据量比较大的情况下，强制是不会进行进行置信区间的计算的

- 默认为不计算
- 还可以选择进行计算，表示会计算 AUC 置信区间，结果会在补充结果部分展示：



AUC信息表

变量	时间(年)	天数	曲线下面积(AUC)	置信区间	累计生存率	累计发病率
group	1	365	0.82298	0.7516 - 0.8944	0.92823	0.07177
group	2	730	0.77483	0.7051 - 0.8445	0.88022	0.11978
group	3	1095	0.83417	0.7855 - 0.8828	0.80133	0.1961
group	4	1460	0.85934	0.8166 - 0.9021	0.75158	0.24842
group	5	1825	0.86228	0.8202 - 0.9044	0.70287	0.29713

· 通过timeROC包计算得到每个时间点对应的曲线下面积和累计生存率

· 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

ROC检验表

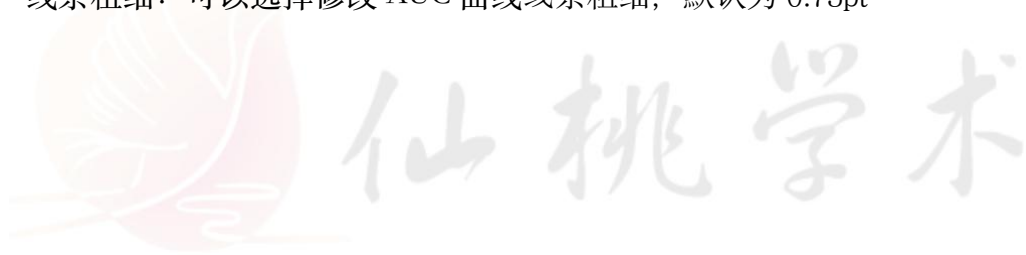
变量1	变量2	时间(年)	天数	p值
group	group1	1	365	8.72e-07
group	group1	2	730	4.93e-06
group	group1	3	1095	6.22e-15
group	group1	4	1460	0
group	group1	5	1825	5.4e-14

· timeROC包提供了检验同一个时间点下不同变量的AUC之间是否有差异的方法(compare)

线



- 颜色：线条颜色，有多少个预测变量取多少个颜色
- 线条类型：AUC 曲线的线条类型，默认为实线，可以选择虚线类型
- 线条粗细：可以选择修改 AUC 曲线线条粗细，默认为 0.75pt



标题文本

标题	
大标题	大标题内容
x轴标题	x轴标题内容
y轴标题	y轴标题内容

- 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]]

图注

图注

是否展示

图注标题

图注标题内容

图注位置

默认

- 是否展示：可以选择是否展示图注信息，默认展示
- 图注标题：可以修改图注标题的内容
- 图注位置：默认为图片的右下，还可以选择右

坐标轴

坐标轴

x轴范围+刻度

逗号隔开

y轴范围+刻度

逗号隔开

- x 轴范围+刻度：可以控制 x 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, 50, 100, 150, 200, 250, 300。(如果调整过大可能会无作用)
- y 轴范围+刻度：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 0.1, 0.1, 0.2, 0.3 (最小值和最大值不能不能可视化数据范围 20%，如果调整过大可能会无作用)

风格



- 边框：是否添加主图的外框
- 网格：是否添加网格
- xy 颠倒：是否进行 x 轴与 y 轴颠倒
- 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

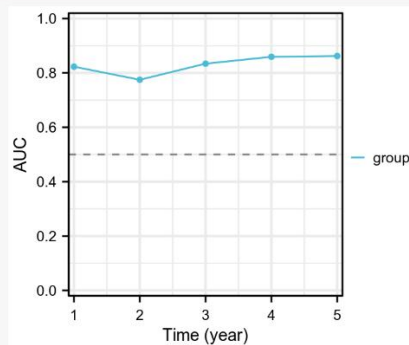


结果说明

主要结果：单个预测变量多个预测时间点

时间依赖AUC

时间依赖AUC: 用折线的形式来展示不同变量在不同时间下其对应曲线下面积 (AUC) 的情况



时间依赖AUC.pdf

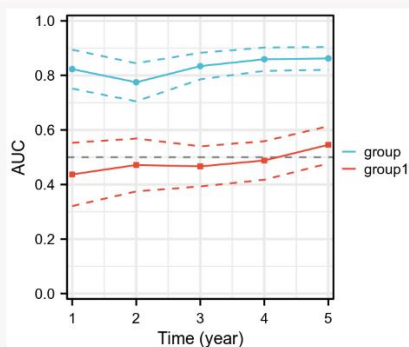
时间依赖AUC.tiff

- 横坐标表示 预后时间; 纵坐标表示 曲线下面积 (AUC)
- 每一条折线表示: 每一个预测变量在参数设定的时间点内的AUC值的变化情况; 虚线表示: 每一个预测变量AUC值的置信区间
- 当一个变量(危险因素)的值是促进事件发生的趋势时, 该分子的AUC会>0.5, 此时面积越大(AUC值越接近于1)说明预后效能越好; AUC在0.5~0.7时, 预后效能较低; AUC在0.7~0.9时, 有一定预后效能; AUC在0.9以上时, 预后效能较好

主要结果：多个预测变量多个预测时间点

时间依赖AUC

时间依赖AUC: 用折线的形式来展示不同变量在不同时间下其对应曲线下面积 (AUC) 的情况



时间依赖AUC.pdf

时间依赖AUC.tiff

- 横坐标表示 预后时间; 纵坐标表示 曲线下面积 (AUC)
- 每一条折线表示: 每一个预测变量在参数设定的时间点内的AUC值的变化情况; 虚线表示: 每一个预测变量AUC值的置信区间
- 当一个变量(危险因素)的值是促进事件发生的趋势时, 该分子的AUC会>0.5, 此时面积越大(AUC值越接近于1)说明预后效能越好; AUC在0.5~0.7时, 预后效能较低; AUC在0.7~0.9时, 有一定预后效能; AUC在0.9以上时, 预后效能较好



补充结果：单个预测变量多个预测时间点的情况

统计描述

变量	时间(年)	天数	事件数(截至当前时间)	生存数(截至当前时间)	删失数(截至当前时间)
group	1	365	30	388	0
group	2	730	50	365	3
group	3	1095	81	312	25
group	4	1460	100	245	73
group	5	1825	115	197	106

· 如果某个时间点的事件数为0，则该时间点无法进行分析和可视化
 · 如果某些时间点的事件数、生存数、删失数相同，建议增加时间间隔（如果时间点大于数据的最大时间，这些结果也会相同）
 · 注意：数据默认在分析前统一删除变量缺失的样本

这里的表格提供不同时间点的事件情况

- 如果某个时间点的事件数为 0，则该时间点无法进行分析和可视化
- 如果某些时间点的事件数、生存数、删失数相同，建议增加时间间隔（如果时间点大于数据的最大时间，这些结果也会相同）
- 数据默认在分析前统一删除变量缺失的样本

ROC结果表

变量	时间(年)	天数	最佳阈值(cut-off)	敏感度	特异度	阳性预测值	阴性预测值
group	1	365	2.3	0.86667	0.69845	0.18182	0.98545
group	2	730	2.2	0.75969	0.69863	0.25542	0.95528
group	3	1095	2.2	0.80302	0.76603	0.45649	0.9408
group	4	1460	2.2	0.78896	0.80408	0.57101	0.92018
group	5	1825	1.9	0.8062	0.80711	0.63858	0.90784

ROC结果表.xlsx

· 先通过（敏感度+(1-特异度)）得到最大对应的第一个 cut_off 作为最佳阈值（cut_off），再通过timeROC包计算得到最佳阈值（cut_off）下的部分ROC相关信息和数据
 · timeROC包提供单个cut_off值的灵敏度、特异度、阳性预测值和阴性预测值的结果，因此无法提供其他ROC相关的指标结果
 · 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算则无法进行相关的时间依赖ROC分析

这里的表格为 ROC 结果表格，提供不同时间点的 ROC 曲线下面积的结果及其其他指标

- 先通过（敏感度+(1-特异度)）得到最大对应的第一个 cut_off 作为最佳阈值（cut_off），再通过 timeROC 包计算得到最佳阈值（cut_off）下的部分 ROC 相关信息和数据

- timeROC 包值提供单个 cut_off 值的灵敏度、特异度、阳性预测值和阴性预测值的结果，因此无法提供其他 ROC 相关的指标结果
- 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

AUC信息表

变量	时间(年)	天数	曲线下面积(AUC)	累计生存率	累计发病率
group	1	365	0.82298	0.92823	0.07177
group	2	730	0.77483	0.88022	0.11978
group	3	1095	0.83417	0.80133	0.1961
group	4	1460	0.85934	0.75158	0.24842
group	5	1825	0.86228	0.70287	0.29713

· 通过timeROC包计算得到每个时间点对应的曲线下面积和累计生存率

· 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

这里的表格为 AUC 信息表，提供不同预测时间点的信息

- 通过 timeROC 包计算得到每个时间点对应的曲线下面积和累计生存率
- 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

补充结果：多个预测变量多个预测时间点的情况（且计算置信区间的前提下）

统计描述

变量	时间(年)	天数	事件数(截至当前时间)	生存数(截至当前时间)	删失数(截至当前时间)
group	1	365	30	388	0
group	2	730	50	365	3
group	3	1095	81	312	25
group	4	1460	100	245	73
group	5	1825	115	197	106
group1	1	365	30	388	0
group1	2	730	50	365	3
group1	3	1095	81	312	25
group1	4	1460	100	245	73
group1	5	1825	115	197	106

- 如果某个时间点的事件数为0，则该时间点无法进行分析和可视化
- 如果某些时间点的事件数、生存数、删失数相同，建议增加时间间隔（如果时间点大于数据的最大时间，这些结果也会相同）
- 注意：数据默认在分析前统一删除变量缺失的样本

ROC结果表

变量	时间(年)	天数	最佳阈值(cut-off)	敏感度	特异度	阳性预测值	阴性预测值
group	1	365	2.3	0.86667	0.69845	0.18182	0.98545
group	2	730	2.2	0.75969	0.69863	0.25542	0.95528
group	3	1095	2.2	0.80302	0.76603	0.45649	0.9408
group	4	1460	2.2	0.78896	0.80408	0.57101	0.92018
group	5	1825	1.9	0.8062	0.80711	0.63858	0.90784
group1	1	365	103.0	0.6	0.23454	0.057143	0.8835
group1	2	730	103.0	0.60013	0.2274	0.095599	0.80691
group1	3	1095	165.0	0.48089	0.39423	0.16267	0.75629
group1	4	1460	165.0	0.45394	0.45306	0.21527	0.71511
group1	5	1825	213.0	0.40636	0.72589	0.38526	0.74309

ROC结果表.xlsx

- 先通过（敏感度+(1-特异度)）得到最大对应的第一个cut_off作为最佳阈值（cut_off），再通过timeROC包计算得到最佳阈值（cut_off）下的部分ROC相关信息和数据
- timeROC包值提供单个cut_off值的灵敏度、特异度、阳性预测值和阴性预测值的结果，因此无法提供其他ROC相关的指标结果
- 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算则无法进行相关的时间依赖ROC分析

AUC信息表

变量	时间(年)	天数	曲线下面积(AUC)	置信区间	累计生存率	累计发病率
group	1	365	0.82298	0.7516 - 0.8944	0.92823	0.07177
group	2	730	0.77483	0.7051 - 0.8445	0.88022	0.11978
group	3	1095	0.83417	0.7855 - 0.8828	0.80133	0.1961
group	4	1460	0.85934	0.8166 - 0.9021	0.75158	0.24842
group	5	1825	0.86228	0.8202 - 0.9044	0.70287	0.29713
group1	1	365	0.43694	0.3209 - 0.553	0.92823	0.07177
group1	2	730	0.47158	0.3747 - 0.5685	0.88022	0.11978
group1	3	1095	0.46642	0.3929 - 0.5399	0.80133	0.1961
group1	4	1460	0.48801	0.4174 - 0.5587	0.75158	0.24842
group1	5	1825	0.5456	0.4775 - 0.6137	0.70287	0.29713

- 通过timeROC包计算得到每个时间点对应的曲线下面积和累计生存率
- 如果对应的时间点内无事件发生（相当于无阳性事件发生），则无法计算对应的曲线下面积

ROC检验表

变量1	变量2	时间(年)	天数	p值
group	group1	1	365	8.72e-07
group	group1	2	730	4.93e-06
group	group1	3	1095	6.22e-15
group	group1	4	1460	0
group	group1	5	1825	5.4e-14

· timeROC包提供了检验同一个时间点下不同变量的AUC之间是否有差异的方法(compare)

- timeROC 包的 compare 函数是可以比较同一个时间点下不同变量的 AUC 之间是否有差异的检验方法
- 当预测变量相同时，其对应的 p 值没有意义，即为 NA

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：timeROC（用于分析），ggplot2 包（用于可视化）

处理过程：

- (1) 通过 timeROC 包对数据进行分析，得到 AUC 相关信息
- (2) 使用 ggplot2 包将所得 AUC 结果进行可视化



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)。

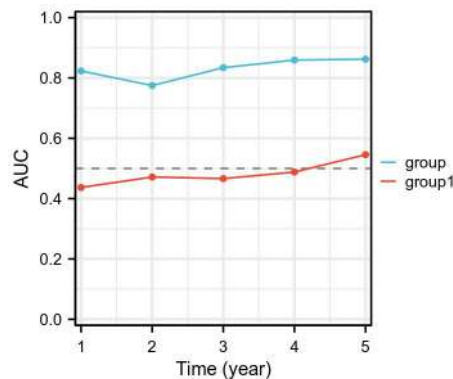
方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。



常见问题

1. 多个预测变量的时候是什么样子的，多个预测变量的时候还能预测多个时间节点吗？

答：多个预测变量结果如图：



预测时间点不变，但是最多只能预测 10 个时间点，最多支持 6 个预测变量

2. 为什么有多个曲线，没有给出统计检验的 p 值？

答：同一个预测变量不同时间点没办法进行统计检验（timeROC 包不提供）；只有不同预测变量同一个时间点才会有统计检验的结果

3. 为什么 AUC 会小于 0.5

答：timeROC 包固定了（无法修改）在一个时间段内发生事件的作为病例组，所以当在一个分组数值的趋势时与预后结局趋势相反（类似保护因素 $HR < 1$ ）时，则会出现 $AUC < 0.5$ ，此时 AUC 越接近 0，说明预测效能越好