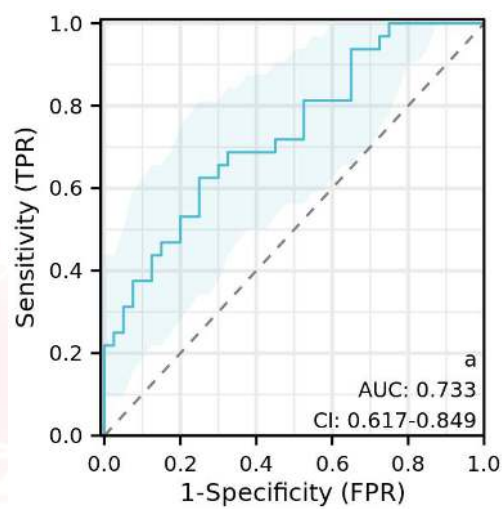


临床意义 - 诊断 ROC 曲线-置信区间



网址: <https://www.xiantao.love>



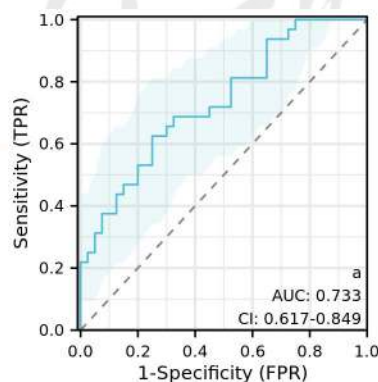
更新时间: 2023.10.08

目录

基本概念	3
应用场景	3
数据格式	5
参数说明	6
数据处理	6
统计	6
线	8
点	10
曲线下面积	错误! 未定义书签。
标题	12
图注	13
风格	14
图片	15
结果说明	16
主要结果	16
补充结果	17
方法学	19
如何引用	20
常见问题	21

基本概念

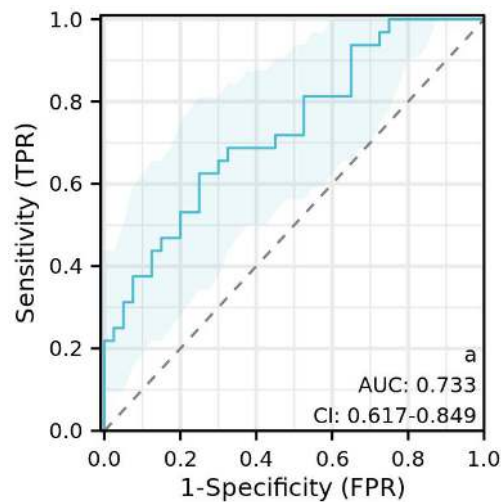
- 诊断 ROC 曲线: 受试者工作特征曲线 (Receiver Operating Characteristic Curve, ROC 曲线) 和 ROC 曲线下的面积 (Area Under ROC Curve, AUC) 常用于诊断试验的评估, 评估预测准确率情况。例如一组数据的结局为 group1 和 group2, 变量为 a、b 和 c, 也就是评估 a、b 和 c 在预测 group1 和 group2 上的结局, 哪个的准确性更高。ROC 曲线图是反映敏感性与特异性之间关系的曲线。AUC 取值范围一般在 0.5 和 1 之间, 使用 AUC 值作为评价标准是因为很多时候 ROC 曲线并不能清晰的说明哪个分类器的效果更好, 而作为一个数值, 对应 AUC 更大的分类器效果更好。
- 诊断 ROC 曲线-置信区间: 在诊断 ROC 曲线的基础上添加曲线下面积(AUC)置信区间
- 图形构成



应用场景

多应用在医学领域, 判断某种因素对于某种疾病的诊断是否有诊断价值。

结果解读



诊断 ROC 曲线

- 横坐标 X 轴为 $1 - \text{特异性}$ ，也称为假阳性率，X 轴越接近零准确率越高；纵坐标 Y 轴称为敏感度，也称为真阳性率，Y 轴越大代表准确率越好。
- AUC (Area Under Curve, AUC)，ROC 曲线下的面积，常用于诊断试验的评估，AUC 取值范围一般在 0.5 和 1 之间，AUC 越接近于 1，说明该变量在预测结局上诊断效果越好。图中 c 变量 (AUC 面积为 0.924) 相比于变量 a 和 b 诊断效果较好。

数据格式

	A	B	C	D
1	outcome	a	b	c
2	group1	1.585854594	1.17428046	2.674787643
3	group1	2.205293427	0.86192791	2.003079333
4	group1	2.199553804	2.31587217	1.281605297
5	group1	1.241118417	1.574637672	1.866428136
6	group1	2.016991788	1.953333649	1.84722131
7	group1	2.391270613	1.089195069	2.149648096
8	group1	0.620790581	0.837543584	1.864922823
9	group1	2.442848378	1.736095106	1.213975139
10	group1	1.636013122	2.414536228	2.946673422
11	group1	1.420847315	2.405175261	1.14110224
12	group1	1.562684913	1.335404124	1.97343619
13	group1	1.034827449	1.373806171	2.070316063
14	group1	2.097985186	2.184310223	2.076315129
15	group1	1.481034166	2.078691615	1.766842006
16	group1	1.020168922	2.388650332	1.862112528

表格 1: 变量预测数据

- 第一列结局变量（必须是二分类），每个结局变量至少需要 2 行数据
- 至少需要 2 列数据，一次最大只支持 11 列数据(10 个自变量)(更多的变量建议是分成 2 张图进行展示)，最少需要 6 行，最多不能超过 10000 行
- 每一行为 1 个样本。
- 除第一列外，其他列必须都是数值（为待分析的变量），填入每个样本对应的变量的值。

例如：如果是表达谱数据，则 abc 代表想要分析的分子，每一行代表 1 个样本，第 1 列代表样本所属于的分组（想要预测的结局，比如正常 vs 异常）

注意：尽量按照示例数据格式整理数据，否则有可能导致验证数据失败。

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**：默认是不处理变量缺失值，可以选择去除任一变量缺失的样本

统计

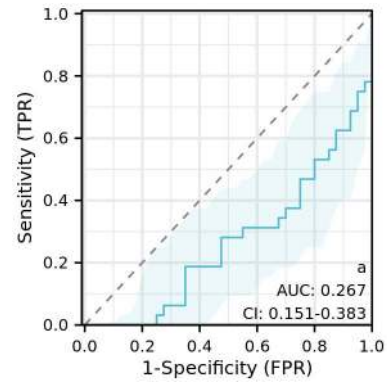


- **方向**：设定二分类结局对应比较的方向(影响图形凹凸方向)，可以选择自动、正向或者反向，如下：

统计

方向 正向

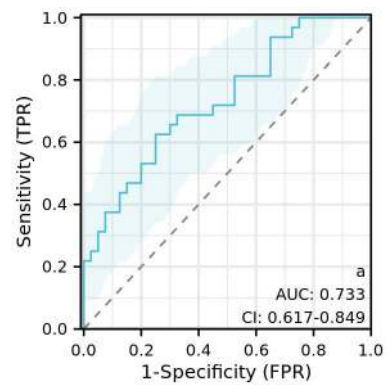
变量 a x



统计

方向 反向

变量 a x



- 变量：选择模型中的需要进行 ROC 可视化的变量，最多支持 4 个，超过之后将会默认不进行置信区间绘制，如下：

统计

方向 自动

变量

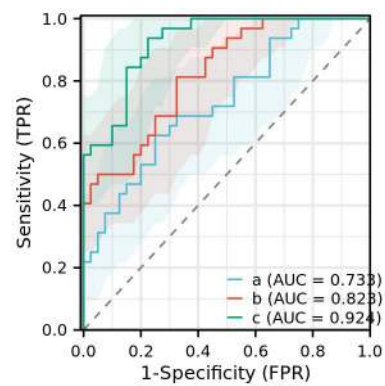
a x

+ 2

a ✓

b ✓

c ✓



置信区间

置信区间

展示

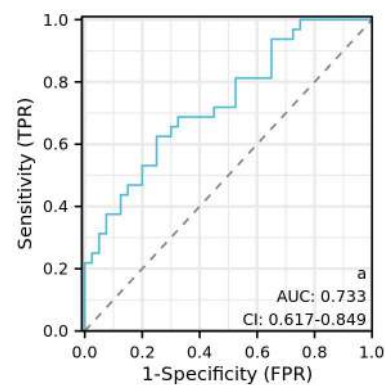
不透明度 0.1

- 展示：控制是否进行置信区间绘制，如下：

置信区间

展示

不透明度 0.1



- 不透明度：当进行置信区间绘制的时候，可以修改置信区间的透明度，1 表示完全不透明，0 表示完全透明

线

线

颜色

样式

实线

粗细

0.75pt

不透明度

1

- 颜色：曲线的颜色
- 样式：默认是实线，也可以选择虚线
- 粗细：默认是 0.75pt
- 不透明度：默认是 1，0 是完全透明，1 是完全不透明

点

点

展示 ☐

填充色

描边色

样式

圆形 ×

大小 0.3

不透明度 1

- 展示：是否展示曲线上的点
- 填充色：点的填充色
- 描边色：点的描边色
- 样式：圆形、三角形等形状选择
- 大小：点的大小，默认 0.3
- 不透明度：默认是 1，0 是完全透明，1 是完全不透明

对角线

对角线

展示

颜色

样式

粗细

虚线

0.75pt

- 展示：是否展示对角线
- 颜色：当进行展示对角线的时候，可以修改对角线的颜色
- 样式：当进行展示对角线的时候，可以修改对角线的样式
- 粗细：当进行展示对角线的时候，可以修改对角线的线条粗细

十字标记

十字标记

展示

颜色

样式

实线

粗细

0.75pt

- 展示：是否展示曲线下面积(AUC)最佳阈值对应特异度和 1-敏感度，默认不进行展示，如下：

十字标记

展示

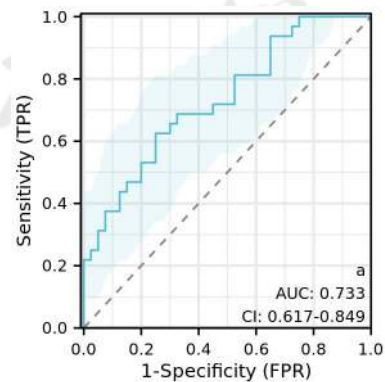
颜色

样式

实线

粗细

0.75pt



十字标记

展示

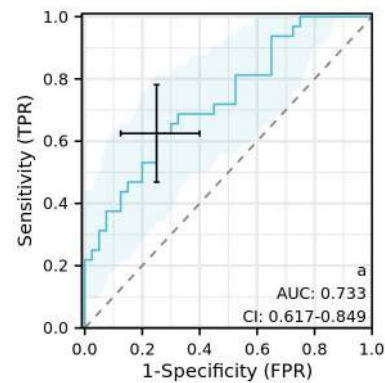
颜色

样式

实线

粗细

0.75pt



标题

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：大标题内容
- x 轴标题：x 轴标题内容
- y 轴标题：y 轴标题内容

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：图注内容是否展示
- 图注标题：可以填入图注标题
- 图注位置：默认是右下，还可以选右

风格

风格

▼

边框

☒

网格

☒

文字大小

7pt

▼

- 边框：是否在图中添加边框
- 网格：是否在图中添加网格线
- 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

图片	▼
宽度 (cm)	5
高度 (cm)	5
字体	Arial ▼

- 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：可以选择图中文本内容字体



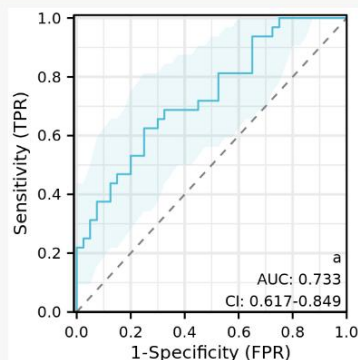
结果说明

主要结果

诊断ROC-置信区间

诊断ROC-置信区间: 用于预测准确率情况

· 预测结局(二分类): group2 vs. group1 (其中参考组: group1) [影响图中灵敏度和特异度以及真假阳性和阴性的计算] [可以通过修改上传数据中的第1列结局的分组出现顺序]



[诊断ROC-置信区间.pdf](#)

[诊断ROC-置信区间.tiff](#)

[诊断ROC-置信区间.pptx](#)

1. ROC曲线图是反映敏感性与特异性之间关系的曲线。横坐标X轴为1-特异性, 也称为假阳性率, X轴越接近零准确率越高; 纵坐标Y轴称为敏感度, 也称为真阳性率(敏感度), Y轴越大代表准确率越好。

2. ROC曲线下的面积 (Area Under Curve, AUC) 常用于诊断试验的评估, AUC取值范围一般在0.5和1之间, AUC越接近于1, 说明该变量在预测结局上诊断效果越好。

主要结果格式为图片格式, 提供 PDF、TIFF、PPT 格式下载。

补充结果

1. 统计描述表：上传数据的一些基本情况

统计描述

各个组常见「统计描述指标」

结局	变量	数目	最小值	最大值	中位数(Median)	四分位距(IQR)	下四分位	上四分位	均值(Mean)	标准差(SD)
group1	a	40	0.58634	2.4428	1.4531	0.82236	1.1199	1.9423	1.5112	0.55454
group1	b	40	0.55563	2.4946	1.8003	1.0719	1.137	2.2089	1.6463	0.61858
group1	c	40	1.0473	2.9467	2.0733	0.66516	1.7701	2.4352	2.0366	0.52699
group2	a	32	0.1287	1.9882	0.99585	0.87824	0.62288	1.5011	0.99804	0.55444
group2	b	32	0.022259	1.9617	0.77933	0.85985	0.32455	1.1844	0.84336	0.57161
group2	c	32	0.030092	1.8936	0.9231	0.96898	0.42143	1.3904	0.93236	0.54622

2. AUC 结果表

AUC结果表

预测变量	预测结局	曲线下面积(AUC)	置信区间(CI)
a	反向	0.733	0.617 - 0.849
b	反向	0.823	0.730 - 0.917
c	反向	0.924	0.868 - 0.981

预测结局中, 正向或者反向会影响真/假阳性和真/假阴性的区分(如果统计-方向参数选择的是“自动”, 则会对结局的方向会进行调整保证曲线都是往上凸(pROC包提供))(如果选择“正向”或者“反向”, 则图形有可能会向下凹)

在AUC > 0.5的情况下, AUC越接近于1, 说明该变量在预测结局上诊断效果越好。

AUC在0.5 ~ 0.7时有较低准确性, AUC在0.7 ~ 0.9时有一定准确性, AUC在0.9以上时有较高准确性。

AUC = 0.5时, 说明该变量不起作用, 无诊断价值。

3. ROC 信息表

ROC信息表

预测变量	cut-off值	灵敏度	特异度	准确率	真阳个数	真阴个数	假阳个数	假阴个数	阳性预测值	阴性预测值	约登
a	1.1155	0.625	0.75	0.69444	20	30	10	12	0.66667	0.71429	0.4
b	1.2633	0.8125	0.675	0.73611	26	27	13	6	0.66667	0.81818	0.4
c	1.7437	0.9375	0.775	0.84722	30	31	9	2	0.76923	0.93939	0.7

各预测变量在各自最佳cut-off值下部分ROC相关信息和数据。

4. AUC 检验表：对变量 AUC 进行检验

AUC检验表

变量1	变量2	统计量	p值	检验方法	趋势方向
a	b	-1.2406	0.2148	DeLong's test	一致
a	c	-2.7917	0.0052	DeLong's test	一致
b	c	-1.7949	0.0727	DeLong's test	一致

1. 变量在AUC > 0.5的情况下，AUC越接近于1，该变量在预测结局上诊断效果越好,具体AUC值参考上面AUC结果表

2. DeLong's test检验结果中若P值小于0.05，则表明两个变量在预测结局上诊断效果具有统计学意义，若大于0.05,则表明两个变量在预测结局上诊断效果没有统计学意义。

3. 两个变量如果趋势方向(direction)不一致，用DeLong方法可能结果会不准确

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. 使用 pROC 包进行对数据进行 ROC 分析，结果用 ggplot2 进行可视化
2. pROC 包默认会对数据的结局顺序进行校正(保证结果是往上凸)



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. AUC 会出现 < 0.5 的情况吗?

答：一般情况下，pROC 分析结果中 AUC 面积是在 0.5-1 之间。

2. 1 个组的时候为什么没有给出统计学检验的 p 值?

答：

ROC 一般是看 AUC 的大小的,只有当存在有多个曲线的时候才会进行检验比较。如果只有 1 条曲线,是没办法进行统计检验的,除非是跟 0.05 的对角线比,这种比较其实是没有意义的,这种只要 AUC 的下限没有跨过 0.5,那么这个曲线肯定是有意义的,所以单个曲线是没有统计学比较的意义。

3. 这里能做联合指标的 ROC 分析吗?

答：

如果上传的数据是 logistic 模型的 predict 值,这个值就可以用来代表联合指标。在 logistic 回归模块或者是诊断列线图模块分析后均会有提供预测值.xlsx 文件下载,从这个里面就可以获取到联合指标的 predict 值,用这个值做 ROC 就是联合指标 ROC 分析。

4. 数据的结局是以哪个作为阴性(参考)? 哪个作为阳性(实验)?

答：

默认上传数据的第一列(二分类)以第一个出现的分类组参考,后出现的分类作为实验。这个方向会影响最终的真阳、真阴、假阳、假阴个数。如果需要反过来,可以在<统计>-<方向>参数中进行修改。



5. 我的数据有很多，但是为什么 统计-变量 没选择几个就会报错？



答：选择模型中的需要进行 ROC 可视化的变量，最多支持 4 个，超过之后将会默认不进行置信区间绘制