

临床预测模型构建&机器学习(R语言进阶)

课程内容概要

周支瑞

课程简介

- R 语言作为一款免费开源的统计软件，已逐渐成为医学科研工作者分析数据的首选软件，很多发表在顶级医学杂志的论文，尤其是临床预测模型构建、数据挖掘类型的论文首选R软件进行统计分析与数据可视化。
- 我们这次推出的《临床预测模型构建&机器学习（R语言进阶）》是之前推出的R语言视频课程《从零开始学 R 语言，带你玩转医学统计学》的进阶版。

课程特点

- 本课程既是医学统计分析进阶培训，也是临床预测模型构建的方法培训，深度讲解临床研究资料统计分析中的新方法与新方法，更包括时髦的机器学习常用方法在处理临床数据及构建预测模型中的应用，课程中的全部案例均为医学相关真实案例；
- 本课程中所有统计分析均基于R语言完成，所以也是一门针对医学专业人员的R语言进阶培训课程！
- 所有案例均提供对应 R语言代码与数据集，可直接学习使用。

什么是临床预测模型？

- 临床预测模型是通过已知参数来预测未知结局，而模型就是一个数学公式。也就是把已知的东西通过这个模型计算未知结局发生的可能性。
- 临床预测模型就是通过回归建模分析，回归的本质就是发现规律。回归是量化刻画，简单讲就是X多大程度上影响Y。尤其是多元线性回归、Logistic回归、Cox回归分析等。也可以通过一些非参数化的方法构建模型，比如机器学习的算法。
- 模型的验证也体现着高度的技术含量。模型效能评价是统计分析、数据建模、课题设计的关键所在。

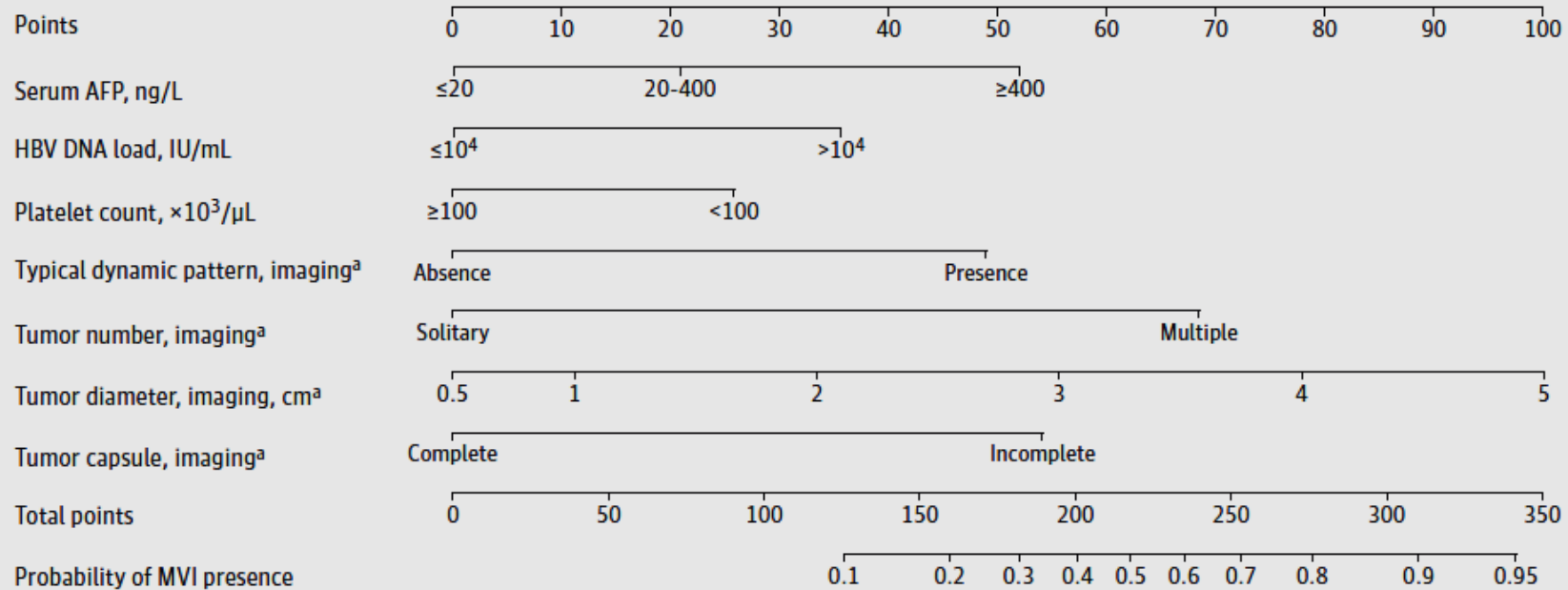
预测模型的呈现形式

- 列线图 Nomogram
- 网页计算器
- 评分系统

列线图

Figure. Nomogram for Preoperative Estimation of Microvascular Invasion (MVI) Risk and Its Predictive Performance

A Nomogram




Estimate the risk
using
**ASSIGN
SCORE**

**Estimate the Risk: NHS
Scotland clinical use**
Enter details in the fields to the
right, then click Calculate.

Failure of the 'calculate' option to
produce the ASSIGN score is most
likely because it is being blocked by
your local security settings. You may
like to check these before notifying
us of a problem.

☐ [Estimate the Risk for
ASSIGN visitors](#)

Name/ Patient ID	<input type="text"/>	notes ▶
Current age	<input type="text"/>	notes ▶
Sex	<input checked="" type="radio"/> M <input type="radio"/> F	notes ▶
Scottish Postcode	<input type="text"/>	notes ▶
Family history of CHD/Stroke	<input checked="" type="radio"/> N <input type="radio"/> Y	notes ▶
Is the subject diabetic?	<input checked="" type="radio"/> N <input type="radio"/> Y	notes ▶
Rheumatoid arthritis?	<input checked="" type="radio"/> N <input type="radio"/> Y	notes ▶
Cigarettes smoked daily or equivalent	<input type="text"/> use mean	notes ▶
Systolic Blood Pressure	<input type="text"/> use mean	notes ▶
Total cholesterol	<input type="text"/> use mean	notes ▶
HDL Cholesterol	<input type="text"/> use mean	notes ▶
<div><div>NEW PATIENT</div><div>PRINT</div><div>CALCULATE</div></div>		



Memorial Sloan-Kettering
Cancer Center

[Prediction Tools](#) ▶ [Breast Cancer Nomograms](#) ▶ [Additional Nodal Metastasis](#)

mskcc.org
PREDICTION TOOLS

Breast Cancer Nomogram: Breast Additional Non SLN Metastases

TEXT SIZE ↓A ↑A

This nomogram can be used to help newly diagnosed breast cancer patients assess the likelihood that their breast cancer has spread to the [sentinel lymph nodes](#).

Enter Your Information Clear Calculate ▶

Frozen Section Performed? ☒ **YES**

Was a frozen section analysis performed during pathological examination? This does not have to be the method that detected the cancer in the sentinel lymph nodes, but it is necessary to know as a variable for this calculator.

Pathological Size (0.1 to 9.0 cm)

Size of the primary tumor, in centimeters.


Tumor Type and Grade

Indicate if tumor type is ductal or lobular, as noted in the pathology report. If ductal, indicate the nuclear grade — I: slight or no variation in the size and shape of the nucleus; II: moderate variation in the size and shape of the nucleus; III: marked variation in the size and shape of the nucleus.

Your Results


[Learn more](#) about your results below.

Probability of Spread to Additional Lymph Nodes **17%**

 **Print These Results**

Make an Appointment

Call us to schedule an appointment or contact us online



评分系统

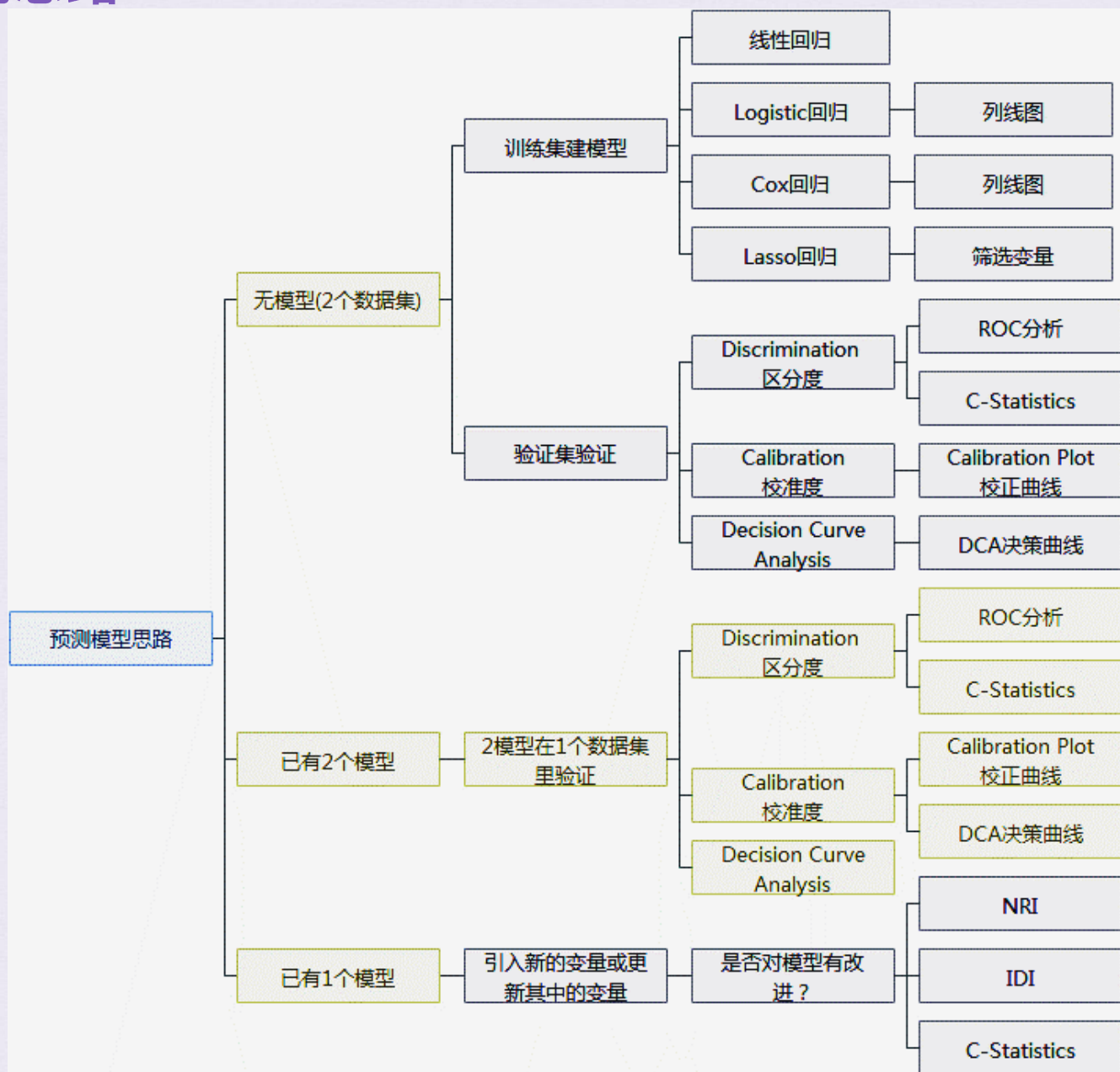
Risk factor	Categories	Points
Age	30-39	0
	40-49	2
	50-59	4
	60-69	6
	70-79	8
Sex	Female	0
	Male	5
Systolic blood pressure	< 120	-1
	120-129	0
	130-139	1
	140-159	2
	≥ 160	3
Current smoker	No	0
	Yes	3

Point total	Estimate of risk	Point total	Estimate of risk
-1	0.0015	10	0.0341
0	0.0020	11	0.0449
1	0.0026	12	0.0590
2	0.0035	13	0.0771
3	0.0047	14	0.1002
4	0.0062	15	0.1293
5	0.0083	16	0.1652
6	0.0110	17	0.2088
7	0.0147	18	0.2602
8	0.0195	19	0.3192
9	0.0258		

临床预测模型构建方式



临床预测模型类研究思路



这门课你将收获

1. 广义线性模型：Logistic回归与泊松回归；
2. Logistic回归变种：含无序多分类Logistic回归，有序Logistic回归等；
3. Logistic回归建模与判别分析；
4. 倾向性匹配得分（PSM）分析；
5. 线性模型中的高级特征选择技术（包括岭回归、Lasso回归、弹性网络）；
6. 生存分析与Cox比例风险模型；
7. Fine&Gray检验与竞争风险模型；
8. 多元回归中变量筛选方法（含亚组分析与敏感性分析）；
9. 回归模型可视化与各类临床预测模型构建（各类回归模型的Nomogram列线图绘制）；
10. 临床预测模型的评价与验证（包括C-statistics计算，NRI与IDI的计算，Calibration曲线绘制，DCA决策曲线分析）；
11. 诊断试验数据处理与ROC分析（包括ROC曲线绘制，AUC计算，联合诊断实现）；
12. 主成分与因子分析、聚类分析在变量降维中的应用；
13. K最近邻与支持向量机在医学统计中应用；
14. 分类回归树（包括回归树、分类树、随机森林等）在医学统计中应用；
15. 神经网络与深度学习在医学统计中应用；
16. 时间序列与因果关系。

适宜人群

1. 本课程主要针对临床医生，医学研究生与临床研究相关专业人员，所有统计分析均是基于R语言，所有案例均为医学研究相关真实案例。
2. 学习本课程应该具备一定的R语言基础知识，强烈推荐在学习本课程之前优先学习我们之前推出的R语言视频课程《从零开始学 R 语言，带你玩转医学统计学》，链接：
<http://class.dxy.cn/clazz/course/206>
3. 对R语言感兴趣的朋友。

感谢观看

THANKS



丁香园特邀讲师 周支瑞