



第 1 章 临床预测模型概述

通过临床预测模型，人们可以运用数理统计学方法，对潜在的自变量 X 进行筛选、整合，构建多变量模型以预测因变量 Y ，并对此模型进行评价、验证。常见的线性回归、Logistic 回归、Cox 回归、竞争风险模型均可以用于构建相应的临床预测模型，其中尤以 Logistic 回归、Cox 回归临床预测模型多见。

临床预测模型的构建有助于医疗卫生人员对临床上某一特定结果（未来）带来的风险进行估计，结合相应的临床经验，较好地进行临床决策。临床预测模型包括诊断模型和预后模型。诊断模型基于研究对象的症状和体征及其他临床资料，判断研究对象是否患某种疾病或处于某种状态。预后模型基于某种疾病或状态，估计研究对象未来死亡、复发某种疾病或其他伤残事件出现的风险。

临床预测模型的结局多为二分类，如是否死亡、是否痊愈，此时可以构建二分类 Logistic 临床预测模型；若结局为等级资料，如心功能分级，可以构建有序 Logistic 临床预测模型；若结局指标包括生存结局、生存时间，可以构建生存资料的临床预测模型，根据生存结局是否包含竞争事件，以决定是拟合 Cox 回归临床预测模型还是拟合竞争风险临床预测模型。除此之外，结局指标还可以是无序多分类或连续性资料。

本章将简要介绍临床预测模型构建、评价、验证的常见方法。

本章主要涉及的知识点：

- 模型构建。
- 模型评价。
- 模型验证。

注意：本章内容仅为概述（读者可结合后续相应章节理解，事半功倍）。

1.1 如何构建预测模型

构建预测模型，关键的步骤是自变量 X 的筛选。目前变量筛选方法较多，各种方法都有其优缺点及适用范围，尚无公认的最优筛选方法。下面对常见的变量筛选方法做简要介绍。

1.1.1 先单后多

最常见的变量筛选方法是先单后多，即先进行单因素分析，再将单因素分析“有意义”的变量一起，纳入多因素模型。这样操作最为简单，并且在大多数情况下是可行的。然而，在某些情况下，先单后多存在一定局限。例如，自变量数目过多，自变量间存在共线性，或者缺失值较多而不愿舍弃掉含有缺失值的样本。

除此之外，先单后多自身也存在一定的争议。例如，对于单因素分析“有意义”的变量，不同人也有不同的理解：有人认为单因素分析 $p < 0.05$ ，即“有意义”；而其他人认为 $p < 0.1$ 或 $p < 0.2$ ，即“有意义”；更甚者认为无论 p 值如何，只要临床上认为其与结局存在关系，便可认为其“有意义”。

注意：本书如未做特殊说明，均认为 $p < 0.05$ “有意义”。

1.1.2 逐步回归

如果在一个方程中，忽略了对因变量 Y 有“显著”影响的自变量，那么所建立的方程必然与实际有较大的偏离。然而，是不是纳入的自变量越多越好呢？显然不是。如果纳入的自变量越多，那么残差平方和 RSS 及其自由度都将减小，但自由度减小的幅度更大，从而使均方误差增大，最终影响预测精度。因此，选择一个“最优”方程十分有必要。

什么是“最优”方程？“最优”方程需要满足两个条件：首先方程能够反映自变量与因变量之间的真实联系，其次方程所使用的自变量数目应尽可能少。

在建立多因素模型时，经常会从影响因变量 Y 的众多因素中挑选部分因素作为自变量建立“最优”模型。此时可以通过逐步回归方法，挑选出合适的自变量。

注意：逐步回归存在一定争议，虽然可能得到一个好的模型，但是不能保证模型是最佳模型，因为不是每一个可能的模型都被评价了。

1.1.3 Lasso

对于高维数据，普通的变量筛选方法并不见效或者需要消耗高昂的计算机算力成本（时间成本）。另外，普通的变量筛选方法也难以避免模型的过度拟合，以及自变量间的多重共线性问题。

通过 Lasso（套索回归），可以对估计出的系数进行限制，避免多重共线性的发生。有时甚至可以将系数缩减至 0，以达到筛选变量的目的。同时由于在残差平方和 RSS 最小化的过程中，加入了正则化项，可以有效地避免过度拟合。

注意：Lasso 可能会将存在共线性的自变量强行剔除。临床上很多指标都会存在共线性。一旦这些指标被剔除，就无法分析其与因变量 Y 的关系。

1.1.4 随机森林

随机森林是有监督的学习方法。在预测模型的构建过程中，通过随机森林可以同时生成多个预测模型，并将模型的结果汇总以提升准确率。

随机森林涉及对样本和自变量进行抽样，从而生成大量的决策树。对每个样本来说，所有决策树依次对其进行预测，预测结果的众数或平均数作为这一样本的最终预测结果。生成决策树未用到的样本也可以由决策树估计，与其真实结果比较，即可得到带外误差。当没有验证数据时，这是随机森林的一大优势。

1.1.5 最优子集

当我们进行模型构建时，通常我们获取到的自变量并不全是有用的，这其中存在着与因变量不相关或者相关性极小的变量。针对这种情况，我们可以根据经验筛选对因变量影响较大的自变量。

然而，通常统计学工作者并不是临床领域的专家，对可能影响因变量的自变量并不了解，于是我们需要运用算法获得最接近真实模型的回归模型，如最优子集回归。

最优子集回归，即对 p 个自变量的所有可能组合分别使用回归进行拟合。总共存在 2^p 个可用于建模的变量子集，根据残差平方和 (RSS) 与 R^2 的改善情况，从中选择一个最优模型。

▶ ▶ 1.2 如何评价预测模型

我们可以从不同角度出发，利用多种指标评价模型。

为什么要进行模型评价呢？因为在模型构建过程中，所拟合的模型不一定是最优模型或者说不是一个好模型，也就是其可能存在欠拟合的情况。常见的评价指标有以下几种。

1.2.1 拟合优度检验

拟合度检验计算每个个体结局事件的预测值，并按照预测值的大小对数据进行分组，

一般 5~10 组，进行 Hosmer-Lemeshow 拟合优度检验，考察预测值与实际值的吻合程度。 $p>0.05$ ，说明模型拟合效果较好。

1.2.2 ROC

人们主要用 ROC (receiver operating characteristic) 曲线来评价预测模型的鉴别能力，即区分能力。ROC 曲线是根据一系列不同的阈值分成两类，以真阳性率 (灵敏度) 为纵坐标，以假阳性率 (1- 特异度) 为横坐标绘制的曲线。ROC 曲线将特异度与灵敏度以图示方法结合在一起，可准确反映模型预测值的特异度与灵敏度的关系。ROC 曲线越接近左上角，曲线下面积越大，表明其预测价值越大。另外，ROC 曲线还可用于不同指标间的比较。一般认为，曲线下面积大于 0.8，其诊断价值较大，具体预测价值需要结合临床实际判断。

1.2.3 Calibration

人们一般使用 Calibration 校准曲线评价预测模型的准确性。Calibration 校准曲线就是以预测发生率为横坐标，以实际发生率为纵坐标的散点图。在散点图上进行直线拟合，如果各点均落在过原点、斜率为 45° 的直线上，则预测模型的准确性非常好；各点离过原点、斜率为 45° 的直线越远，则预测模型的准确性越差。

1.2.4 DCA 曲线

大多数预测模型不能达到 100% 准确，无论选取哪个值作为阈值，都会遇到假阳性和假阴性的问题，有时候避免假阳性会受益更大，有时候则更希望可以避免假阴性。以上两种情况都无法避免，要找到一个净受益最大的办法 (预测模型的实用性问题)，一般采用 DCA (decision curve analysis) 曲线评价预测模型。DCA 曲线以阈概率为横坐标，以利减去弊之后的净获益率为纵坐标。曲线整体越靠近右上角，表明预测模型的实用性越好。在实际应用时还有两条参考线，它们代表两种极端情况。横线表示所有样本都是阴性的，所有人都未接受干预，净获益为 0。斜线表示所有样本都是阳性的，所有人都接受了干预，净获益曲线是一个斜率为负值的反斜线。如果 DCA 曲线靠近这两条线，则其临床实用性较差。

▶▶ 1.3 如何验证预测模型

通常来讲，完成模型评价后的研究已经可以被称为“完整”的研究。然而，很多人会忽略过度拟合这种情况，也就是结果的外推性究竟如何。什么是过度拟合？简单地讲，过度拟

合造成所构建的模型对本次数据显示出很好的效果，但是将这个模型用于预测另外一份全新的数据时，可能呈现出效果不理想的状态。模型验证的方法非常多，这里罗列常用的几种。

1.3.1 交叉验证

所谓交叉验证，即将一定比例的数据挑选出来作为训练集，将其余未选中的样本作为测试集，先在训练集上构建模型，再在测试集上做预测。由于测试集不涉及参数的选择，其可以评估模型的泛化能力。

交叉验证可以按照不同的方法进行分类。最常见的就是交叉验证分为内部验证、外部验证。内部验证即手动将样本随机分成训练集、测试集，先在训练集上获取模型，再在测试集上进行测试。由于测试集不涉及参数的选择，其可以获得更为保守的估计值。外部验证即基于内部数据建模完成之后，其他独立的研究团队开展相同研究，其数据集作为外部数据的来源，进行模型的验证。相对来说，外部验证在统计学方法上易于实现，但资料的获取比较困难。

内部验证可以根据其形式分成以下 3 种：① Hold-Out Method，姑且将其翻译成简单交叉验证；② K -fold Cross Validation（记为 K -CV），即 K 折交叉验证；③ Leave-One-Out Cross Validation（记为 LOO-CV），即留一法交叉验证，也称为 N 折交叉验证。

注意：简单交叉验证最为常用，但从严格意义上来讲，简单交叉验证并不是交叉验证。

1.3.2 Bootstrap 法

Bootstrap 法是非常有用的非参数统计方法，实质是对样本进行重抽样，从而估计总体特征。因为 Bootstrap 法充分利用了已知的样本信息，不需要其他假设或增加新样本，具有一定的稳健性。另外，Bootstrap 法可以避免交叉验证造成的样本减少问题，也可以用于创造样本随机性。

▶ ▶ 1.4 小结

需要说明的是，模型构建、模型评价、模型验证可能需要多次操作，才能得到最终的结果。例如，在模型评价阶段，发现模型存在欠拟合状况，或者在模型验证阶段，发现模型存在过拟合状况，这时候都需要返回模型构建阶段重新拟合模型。