R语言:列线图(Nomogram)的绘制及解读结果

作者：医小咖 来源：医咖会

本文原地址：https://www.iikx.com/news/statistics/11906.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

手把手教你用R画列线图(Nomogram)及解读结果。

同样是构建多因素回归模型，往往我们另一个主要目的是为了对结局事件的发生风险进行预测，那么是否也可以将预测模型的结果，像森林图那样可视化地展示出来呢?本文教你用R画列线图(Nomogram)及解读结果。

认识列线图.

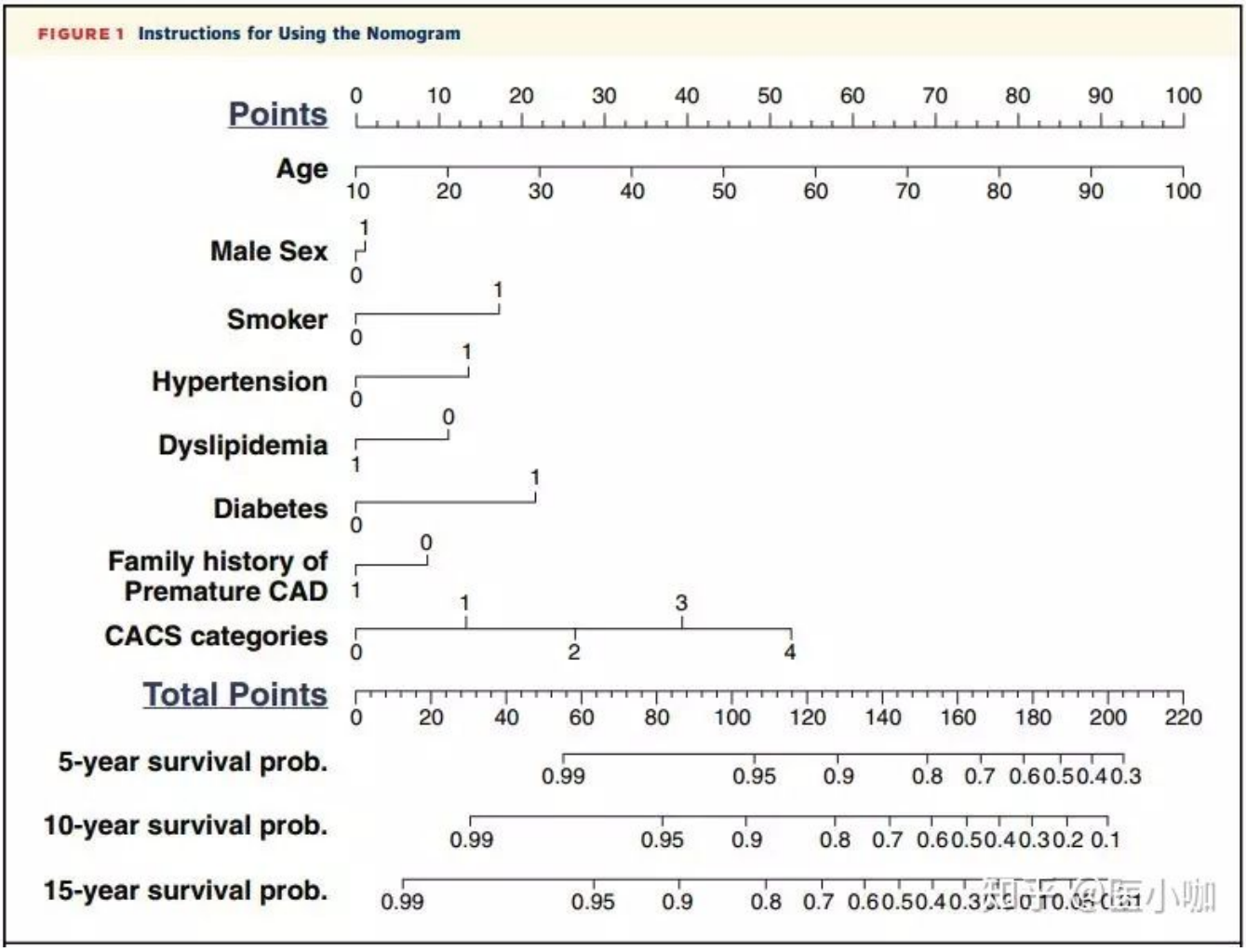
列线图(Alignment Diagram)，又称诺莫图**(Nomogram)**，它是建立在多因素回归分析的基础上，将多个预测指标进行整合，然后采用带有刻度的线段，按照一定的比例绘制在同一平面上，从而用以表达预测模型中各个变量之间的相互关系。

列线图的基本原理，简单的说，就是通过构建多因素回归模型(常用的回归模型，例如Cox回归、Logistic回归等)，根据模型中各个影响因素对结局变量的贡献程度(回归系数的大小)，给每个影响因素的每个取值水平进行赋分，然后再将各个评分相加得到总评分，最后通过总评分与结局事件发生概率之间的函数转换关系，从而计算出该个体结局事件的预测值。

列线图将复杂的回归方程，转变为了可视化的图形，使预测模型的结果更具有可读性，方便对患者进行评估。正是由于列线图这种直观便于理解的特点，使它在医学研究和临床实践中也逐渐得到了越来越多的关注和应用。

解读列线图.

闲话少说，先上文章哈。今天我们以2017年发表在JACC：Cardiovascular Imaging杂志上的一篇文章《Development and Validation of a Simple-to-Use Nomogram for Predicting 5-, 10-, and 15-Year Survival in Asymptomatic Adults Undergoing Coronary Artery Calcium Scoring》为例来进行说明，文中结果部分展示的列线图如下图所示。



我们可以看到列线图主要由左边的名称以及右边对应的带有刻度的线段所组成。

列线图的名称主要包括三类：

1. 预测模型中的变量名称：例如图中的年龄(Age)、高血压(Hypertension)、糖尿病(Diabetes)等信息，每一个变量对应的线段上都标注了刻度，代表了该变量的可取值范围，而线段的长度则反映了该因素对结局事件的贡献大小。

2. 得分，包括单项得分，即图中的Point，表示每个变量在不同取值下所对应的单项分数，以及总得分，即Total Point，表示所有变量取值后对应的单项分数加起来合计的总得分。

3. 预测概率：例如图中的5-year survival prob，表示5年的生存概率。

介绍了列线图的基本要素，下面小咖来教大家如何利用这张列线图，让自己也能成为一个神奇的预言家。敲黑板，讲到重点了哈。

现在假设我们是一名优秀的心内科医生，有这样一位患者，男性，60岁，吸烟，有高血压和糖尿病史，血脂异常，否认CAD家族史，冠状动脉钙化评分(CACS)为3分。

小咖作为接诊医生，在和患者交代病情的时候，为了向患者说明疾病的严重性，就拿出了这张列线图，自信满满的告诉这位患者，以他目前的疾病状态，预测未来5年、10年和15年的生存概率分别是71%、48%和27%。

那么，小咖是怎么算出来的呢?其实很简单，比如该患者年龄为60岁，我们就在列线图年龄为60岁的地方向上画一条垂直线，即可得到其对应的得分(Points)约为55分。同样性别为男性，对应的分数为1分，以此类推，找出每个变量状态下对应的得分。

最后将所有变量的得分相加，得到患者的总得分(Total Points)约为165.5分，并以总得分为基础，再向下画一条垂直线，就可以知道该患者对应的未来5年、10年和15年的生存率了，是不是很简单很容易理解呢!

变量	状态	得分
年龄	60 岁	55
性别	男性（=1）	1
吸烟	是（=1）	17.5
高血压	是（=1）	13
血脂异常	是（=1）	0
糖尿病	是（=1）	22
CAD 家族史	否（=0）	18
CACS	3 分	39
总分		165.5

知乎 @医小咖

列线图的效果评价.

当然，制作列线图通常要求研究拥有足够数量的研究对象，从而才能够建立有效的预测模型，在列线图制作之前也需要对预测模型的预测能力进行验证。常用的效果评价方式有：

1. 内部验证法.

可采用Bootstrap自抽样法，利用建模自身的数据来验证模型的预测效果。Bootstrap自抽样法是指对样本人群进行有放回的重复抽样，每次抽样样本数相同，这样同一个个体就有可能被抽中多次

。利用Bootstrap自抽样产生的新样本去评价列线图模型的准确性，常用C-统计量来进行衡量，其值越接近于1说明列线图的预测能力越准确。

2. 图形校准法.

图形校准法的基本思想是：首先利用列线图预测出每位研究对象的生存概率，并从低到高排成一个队列，根据四分位数将队列分为4组(或者根据其他分位数分组)，然后分别计算每组研究对象预测生存概率和相应的实际生存概率(由Kaplan-Meier法计算)的均值，并将两者结合起来作图得到4个校准点，最后将4个校准点连接起来得到预测校准曲线。

理论上标准曲线是一条通过坐标轴原点、且斜率为1的直线，如果预测校准曲线越贴近标准曲线，则说明列线图的预测能力越好。

3. 外部验证法.

使用一组研究对象去建立列线图，再使用另外一组研究对象(即外部数据)来验证列线图预测效果的准确性。

绘制列线图.

其实，列线图的绘制也很简单，R软件中的rms程序包，为我们提供了相应的函数功能，最后给大家安利一下绘制列线图最基本的程序代码，以供有兴趣的小伙伴进行学习和参考哈。

1.加载survival和rms程序包，以survival包中自带的数据库lung为例.

```
library(survival)

library(rms)

data(package="survival")
```

Filter

	inst	time	status	age	sex	ph.ecog	ph.karno	pat.karno	meal.cal	wt.loss
1	3	306	2	74	1	1	90	100	1175	NA
2	3	455	2	68	1	0	90	90	1225	15
3	3	1010	1	56	1	0	90	90	NA	15
4	5	210	2	57	1	1	90	60	1150	11
5	1	883	2	60	1	0	100	90	NA	0
6	12	1022	1	74	1	1	50	80	513	0
7	7	310	2	68	2	2	70	60	384	10
8	11	361	2	71	2	2	60	80	538	1
9	1	218	2	53	1	1	70	80	521	16
10	7	166	2	61	1	2	70	70	271	34

知乎 @医小咖

(数据格式).

2.打包数据.

dd<-datadist(lung)

options(datadist="dd")

3.构建Cox比例风险回归模型.

```
f <- cph(Surv(time,status) ~ age+sex+ph.karno, data = lung,
         x=TRUE,y=TRUE,surv=TRUE)
```

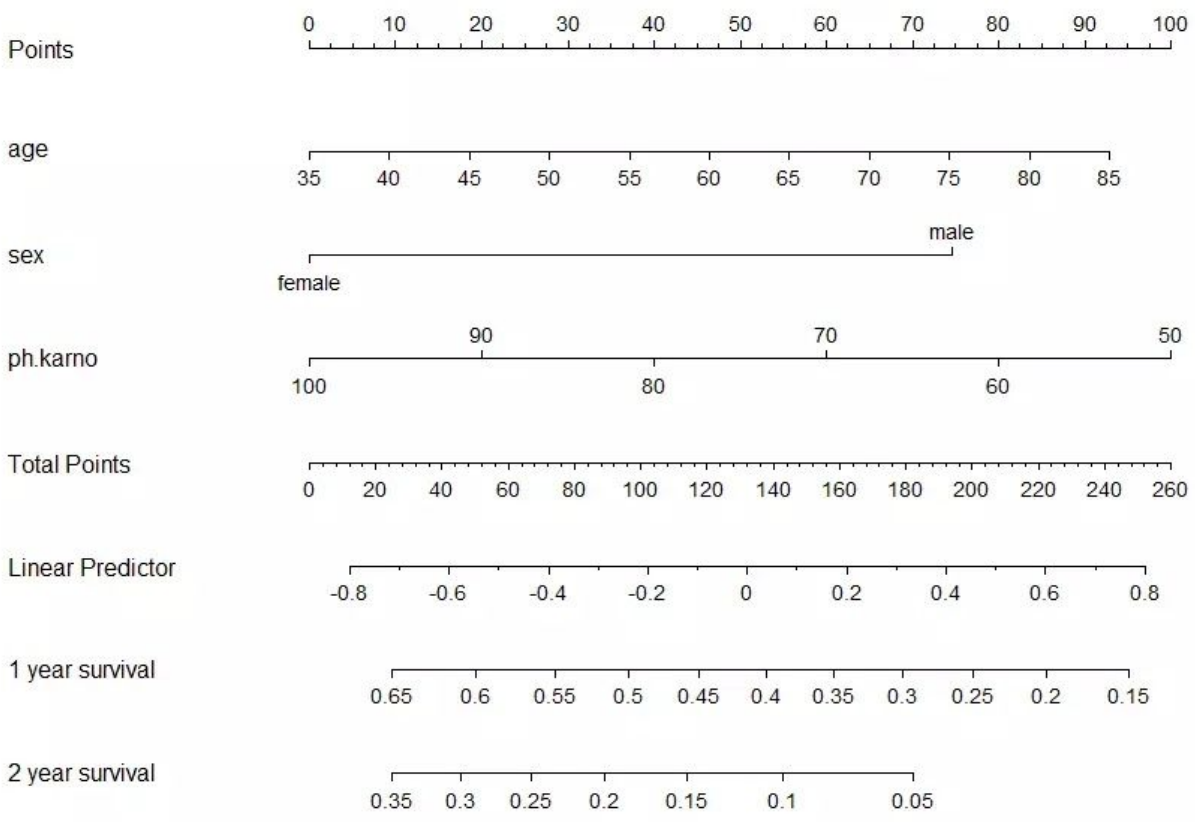
4.绘制列线图.

```
survival<-Survival(f)
survival1<-function(x)survival(365,x)
survival2<-function(x)survival(730,x)

nom<-nomogram(f,fun = list(survival1,survival2),
              fun.at = c(0.05,seq(0.1,0.9,by=0.05),0.95),
              funlabel = c('1 year survival','2 year survival'))
plot(nom)
```

知乎 @医小咖

最终绘制的列线图如下：



知乎 @医小咖

当我们在研究中习惯了使用相对风险的指标时(例如OR、HR、RR等)，也不妨将这种能够起到预测生存概率作用的列线图结合起来，这样会使预测模型的结果更加形象、直观、易懂。

患者可以根据自身各个影响因素的水平大小，能够快速查询到自己未来的生存概率，便于引起他们对健康状况的重视。

更多 统计方法 请访问 https://www.iikx.com/news/statistics/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发