



中山大學
SUN YAT-SEN UNIVERSITY

本科生毕业论文（设计）

Undergraduate Graduation Thesis (Design)

题目 Title: 探究特定亚群下
年龄对乳腺癌患者的影响

院系
School (Department): 数学学院

专业
Major: 统计学

学生姓名
Student Name: 陈敬贤

学号
Student No.: 15338013

指导教师（职称）
Supervisor (Title): 尤娜 教授

时间：二〇一九年五月一日

Date: May 1st 2019

表一 毕业论文（设计）开题报告

Form 1: Research Proposal of Graduation Thesis (Design)

论文(设计)题目:

Thesis (Design) Title: 探究特定亚群下年龄对乳腺癌患者的影响

(简述选题的目的、思路、方法、相关支持条件及进度安排等)

(Please briefly state the research objective, research methodology, research procedure and research schedule in this part.)

选题目的:

乳腺癌是女性中最常被诊断出的癌症，本研究的目的是为了更精确地评估年龄预后因素对小乳癌患者的影响，为临床应用提供帮助。

思路:

本研究通过对患者特征划分亚群，探究年龄对它们生存风险影响的差异，并总结年龄预后因素对患者的影响，提高临床治疗应用的价值。

方法:

本研究首先应用决策树方法划分亚群，然后用 COX 比例风险模型对每个亚群中的患者进行建模，以 c-index 评估模型预测精度，并利用模型年龄预后因素的系数比较它们对生存风险影响的差异。

相关支持条件:

国内外的相关研究成果，来自 REDCap 的 1239 例乳腺癌患者数据。

进度安排:

2019 年 1 月定题；2019 年 2-3 月对数据进行分心并完成程序实现；2019 年 4 月 1-15 日完成论文初稿写作；2019 年 4 月 16-27 日完成论文终稿写作。

Student Signature:

Date:

指导教师意见:

Comments from Supervisor:

1. 同意开题 2. 修改后开题 3. 重新开题
1. Approved() 2. Approved after Revision() 3. Disapproved()

Supervisor Signature:

Date:

表二 毕业论文（设计）过程检查记录表

Form 2: Process Check-up Form

指导教师分阶段检查论文的进展情况（要求过程检查记录不少于 3 次）

The supervisor should check up the working process for the thesis (design) and fill up the following check-up log. At least three times of the check-up should be done and kept on the log.

第一次检查（First Check-up）：

学生总结

Student Self-Summary:

在这一阶段，针对本研究的程序设计，修改工作基本完成

指导教师意见

Comments of Supervisor:

第二次检查（Second Check-up）：

学生总结

Student Self-Summary:

应用实际数据后得到的结果基本呈现，包括划分亚群的决策树，各亚群 COX 比例风险模型的一致性指数和年龄预后因素系数比较。

指导教师意见

Comments of Supervisor:

...

第三次检查 (Third Check-up) :

学生总结

Student Self-Summary:

参考国内外相关文献的结果，结合本研究的实际数据应用结果进行分析总结。

指导教师意见

Comments of Supervisor:

...

第四次检查 (Fourth Check-up) :

学生总结

Student Self-Summary:

对论文的引言，问题应用，结果讨论等部分进行修改，补充并完善。

指导教师意见

Comments of Supervisor:

...

学生签名 (Student Signature) :

日期 (Date) :

指导教师签名 (Supervisor Signature) :

日期 (Date) :

总体完成情况

(Overall Assessment)

指导教师意见 Comments from Supervisor:

- 1、按计划完成，完成情况优 (Excellent): ()
- 2、按计划完成，完成情况良 (Good): ()
- 3、基本按计划完成，完成情况合格 (Fair): ()
- 4、完成情况不合格 (Poor): ()

指导教师签名 (Supervisor Signature) :

日期 (Date) :

表三 毕业论文（设计）答辩情况登记表

Form 3: Thesis Defense Performance Form

答辩人 Student Name	陈敬贤 Chen Jingxian	专业 Major	统计学 Statistics
论文(设计)题目 Thesis (Design) Title	探究特定亚群下年龄对乳腺癌患者的影响 Effect of age on survival in specific subgroups of small breast cancer patients		
答辩小组成员 Committee Members			

答辩记录
Records of Defense Performance:

记录人签名 (Clerk Signature) : 日期 (Date) :

学术诚信声明

本人所呈交的毕业论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。本毕业论文的知识产权归属于培养单位。本人完全意识到本声明的法律结果由本人承担。

本人签名：陈敬贤 日期：2019 年 9 月 29 日

Statement of Academic Integrity

I hereby acknowledge that the thesis submitted is a product of my own independent research under the supervision of my supervisor, and that all the data, statistics, pictures and materials are reliable and trustworthy, and that all the previous research and sources are appropriately marked in the thesis, and that the intellectual property of the thesis belongs to the school. I am fully aware of the legal effect of this statement.

Student Signature: Chen Jingxian

Date: 09-29-2019

探究特定亚群下年龄对乳腺癌患者的影响

[摘要] 背景：乳腺癌作为当前社会的重大公共卫生问题，随着对其认识的不断深入以及治疗理念的转变与更新，乳腺癌的治疗进入了综合治疗时代。因此对于不同特征的病人作出准确的评估，针对地采取更合适的治疗手段，避免无效治疗，从而提高患者的生存率和生活水平，成为当前的一大热点。方法：本研究的目的是找出特定的乳腺癌患者亚群，使得年龄因素对这些特定亚群预后影响的差异明显展示，并提供预测精度更高的模型，进而为临床医治提供有利的参考。本研究对 1239 例乳腺癌病例进行分析，应用决策树模型对患者进行切分亚群，使得每个亚群中年龄对患者的生存影响存在明显差异，并且应用 COX 风险模型来量化评估年龄因素在各亚群中的差异大小及根据一致性指数衡量模型的预测精度。成果：运用决策树根据预后因素 Ki67、Her2 及年龄划分了 4 个亚群；其中亚群 1 和 2 年龄对患者生存风险具有增加作用，并且作用效果大小和 Her2 表达有关；亚群 3 和 4 年龄对患者生存风险具有抑制作用，并且作用效果大小和 Ki-67 蛋白质表达有关；划分亚群后的模型预测精度有明显提高。结论：在预后因素 Ki67、Her2 和年龄所决定出的 4 个特定亚群中，年龄预后因素对乳腺癌患者的影响具有明显差异，划分亚群后模型的预测精度提高，更加有临床应用价值。

[关键词] 乳腺癌；COX 比例风险模型；年龄预后因素；决策树；一致性指数

Effect of age on survival in specific subgroups of small breast cancer patients

[Abstract] Background: as a major public health issue in the current society, the treatment of breast cancer has entered the era of comprehensive treatment with the deepening of its understanding and the transformation and update of the treatment concept. Therefore, it has become a hot topic to make accurate assessment of patients with different characteristics, adopt more appropriate treatment methods and avoid ineffective treatment, so as to improve the survival rate and living standard of patients.

Methods: This study analyzed 1239 cases breast cancer patients and built up a decision tree model to partition subgroups of patients, making each subgroup in the largest impact on survival in patients with age difference. Also we applied COX model to quantify risk assessment age factors in each subgroup and measured accuracy of the model prediction according to the concordance index.

Results: four subgroups were divided according to prognostic factors Ki67, Her2 and age by decision tree. The age of subgroups 1 and 2 promoted the survival risk of patients. The age of subgroups 3 and 4 inhibited the survival risk of patients. The prediction accuracy of the model is improved obviously after subgroup division.

Conclusion: among the 4 specific subgroups determined by the prognostic factors Ki67, Her2 and age, the influence of age prognostic factors on breast cancer patients is significantly different, and the prediction accuracy of the subgroup model is improved, which has more clinical application value.

[Keywords] Breast cancer, COX proportional hazard model, Age prognostic factor, Decision tree, Concordance index

目录

第一章	引言	1
1.1	选题背景与意义	1
1.2	本文的论文结构与章节安排	2
第二章	数据分析	3
2.1	数据介绍	3
2.2	数据基本统计描述	4
2.3	缺失数据的处理	4
第三章	知识点和方法	5
3.1	方法介绍	5
3.2	统计学知识点描述	6
第四章	实验与结果	10
4.1	结果呈现	10
4.2	结果分析	13
4.3	结果讨论	14
第五章	总结与展望	16
5.1	工作总结	16
	参考文献	17
	致谢	19
附录 A	supplement	20
A.1	各个模型的具体参数展示	20

第一章 引言

1.1 选题背景与意义

乳腺癌是女性中最常被诊断出的癌症，如今是一个经常被讨论的问题。从医学的角度来说，识别早期的乳腺癌十分重要但也十分困难，早期发现的乳腺癌是可以治疗的，但仍然有 30% 左右的乳腺癌女性患者死于该癌症^[1]，识别乳腺癌中对治疗效益有意义的预后因素已经成为了近年来研究中很重要的一个主题，评估年龄因素对小乳癌患者的复发风险是正是其中一个热点。当应用一些临床实验的发现时，结论往往是和对患者总体的影响相关的。鉴于我们付出了大量的资源与精力来进行临床实验，除了关于患者总体的结论之外，还能不能得到更多的信息，成为了一个很重要的问题。比如说是否可以获取基本的信息来评估某种预后因素的有效性或者来描述某种预后因素在患者亚群中的异质性，甚至在患者个体中的异质性呢？想要回答类似的问题，进行探索性的分析不可或缺。一些研究把这一类的分析称作“亚群分析”，并认为需要谨慎进行分析由于可能会夸大分析结论和产生误导的结果^[2]。但无论如何，这一类探索性的分析是一种有用的方法来提供合理的假设。

年龄作为乳腺癌的一个重要的风险因素，在大于 50 岁的妇女当中，大约有 78% 被诊断有乳腺癌，87% 的死亡和乳腺癌相关^[3]。然而，在全世界范围内乳腺癌患者数量一致呈不断升高的趋势，尤其是年轻的乳腺癌患者发病率不断上升，使得 40 岁以下的妇女中乳腺癌成为了最常被诊断的癌症^[4]。因此了解清楚诊断年龄和乳腺癌生存之间的关系十分关键。尽管有许多研究推断年龄和乳腺癌的生存有关，但它们之间的关系仍有存在着一定的争议：目前许多患者数据表明年龄更小会和预后不良有关，因为这一群体会存在更多的侵袭性疾病^[5]；但同样有研究表明老年妇女患者的预后会比年轻妇女更差^[6]。周春阳等人^[7]曾分析过小于 36 岁的青年组乳腺癌患者的术后复发率会比 36 岁以上的老年组患者高；Zhong, W et al.^[8]通过运用 PSM 分析方法描述了 40 岁以下妇女患者比 40 岁以上的具有更高的术后复发风险；Rosemary et al.^[9]专门探究的是 55 岁以上的乳腺癌病人的情况；回顾这些研究，我们发现大部分使用不同的年龄切分点和年龄段范围以及关注较小的数据集，这样或许能够解释他们观测到的结果的矛盾。因此，年龄和乳腺癌预后的关系还是不清楚也存在争议。

在本文中，我们描述了一种迭代划分技术^[10]来识别出多种预后因素对生存影响差

异最大的病人分组，并描述了年龄预后因素在各个划分出的患者亚群当中对患者生存影响的差异以及刻画了每个患者亚群中建立模型的预测精度。根据这些特定亚群的划分特征，我们就能够提供一种较客观合理的年龄切分点和年龄段范围。同时保证数据集足够大，我们使用来自 REDCap 总共 1239 例乳腺癌患者的来进行研究和评估我们的方法。据此，本研究能够帮助临床应用中特定的病理特征的患者来选取适合的预后治疗手段。

1.2 本文的论文结构与章节安排

本文共分为五章，各章节内容安排如下：

第一章引言。

第二章数据分析。

第三章知识点和方法。

第四章讨论和结果。

第五章是本文的最后一章, 总结与展望。

第二章 数据分析

2.1 数据介绍

本研究所使用的乳腺癌患者数据来自 REDCap, 总共 1239 例患者数据, 共 11 个变量, 包括: time, status, age, tumor.size, histological.type, tumor.grade, ER, PR, HER2, Ki67, surgery.type.

各变量的含义如下:

1. time: 从开始记录到患者复发或删除的时间;
2. status: 0 代表患者在 $t = time$ 时间点发生删失; 1 代表患者在该时间点发生复发;
3. age: 患者开始被记录时的年龄;
4. tumor.size: 患者的乳腺癌肿瘤大小;
5. ER: 雌激素受体阴阳性 (0 或 1);
6. PR: 孕激素受体阴阳性 (0 或 1);
7. HER2: 表皮生长因子受体阴阳性 (0 或 1);
8. Ki67: Ki-67 蛋白质的表达的两种形态 (0 或 1);
9. surgery.type:

治疗手段的 6 种类型: 双侧乳腺切除术 (双乳癌); 改良乳腺癌根治术 (切乳房 + 腋窝清扫 + 可能部分胸大/小肌); 保乳手术 (包括区段、微创); 单纯乳腺切除术 (切乳房 + 可能腋窝前哨); 根治性乳腺切除 (切乳房 + 腋窝清扫 + 胸大肌 + 胸小肌); 保留乳头的乳腺切除术 (为了乳腺重建)

10. histological.type:

癌症组织形态的 7 种类型: 伴神经内分泌特征的癌 (未分类, 含小细胞癌); 伴髓样特征的癌 (未分类); 浸润性癌 (未分类, 含单纯癌); 浸润性导管癌; 浸润性乳头状癌; 浸润性微乳头状癌; 浸润性小叶癌 (含经典型小叶癌、实性经典型小叶癌)

11. tumor.grade: 三种乳腺癌肿瘤的类型 (1, 2, 3)

2.2 数据基本统计描述

下面的表格展示的是使用的生存数据的基本描述，包括生存时间，状态在内，共 11 个变量；交待了各个变量非缺失值的数量，均值，标准差，中位数，最小值，最大值，范围等基本描述性统计量：

	vars	n	mean	sd	median	min	max	range	se
time	1	1239.00	1746.05	1291.09	1423.00	0.00	5873.00	5873.00	36.68
status	2	1239.00	0.08	0.26	0.00	0.00	1.00	1.00	0.01
age	3	1239.00	50.55	10.97	49.57	21.31	86.87	65.56	0.31
tumor.size	4	1233.00	1.37	0.57	1.50	0.00	3.00	3.00	0.02
histological.type*	5	1239.00	4.09	0.67	4.00	1.00	7.00	6.00	0.02
tumor.grade*	6	809.00	2.22	0.64	2.00	1.00	3.00	2.00	0.02
ER*	7	1225.00	1.84	0.36	2.00	1.00	2.00	1.00	0.01
PR*	8	1225.00	1.75	0.43	2.00	1.00	2.00	1.00	0.01
Her2*	9	1096.00	1.23	0.42	1.00	1.00	2.00	1.00	0.01
Ki67*	10	1089.00	1.75	0.43	2.00	1.00	2.00	1.00	0.01
surgery.type*	11	1239.00	1.89	1.30	1.00	1.00	6.00	5.00	0.04

2.3 缺失数据的处理

生存数据当中一个很常见的问题是患者的部分数据缺失。因此，为了解决建模过程中的这一问题，存在缺失数据的患者需要进行特殊的处理。

有很多不同的方法被提出用来解决这一问题^[11]，最简单的一种是直接排除掉这部分缺失数据的患者，但这样会引起一大部分的患者数据被拒绝。另外一种常见方法是只去掉预后因素中缺失数据的部分，但这种方法很难控制，因为若很多重要的预后因素都被剔除的话有可能会造成模型作出比较差的预测。由于本研究所使用的 1239 例乳腺癌患者数据中缺失值并不多，只有 148 例数据有部分缺失，因此我们直接在部分建模分析时考虑舍弃这小部分的数据。

第三章 知识点和方法

3.1 方法介绍

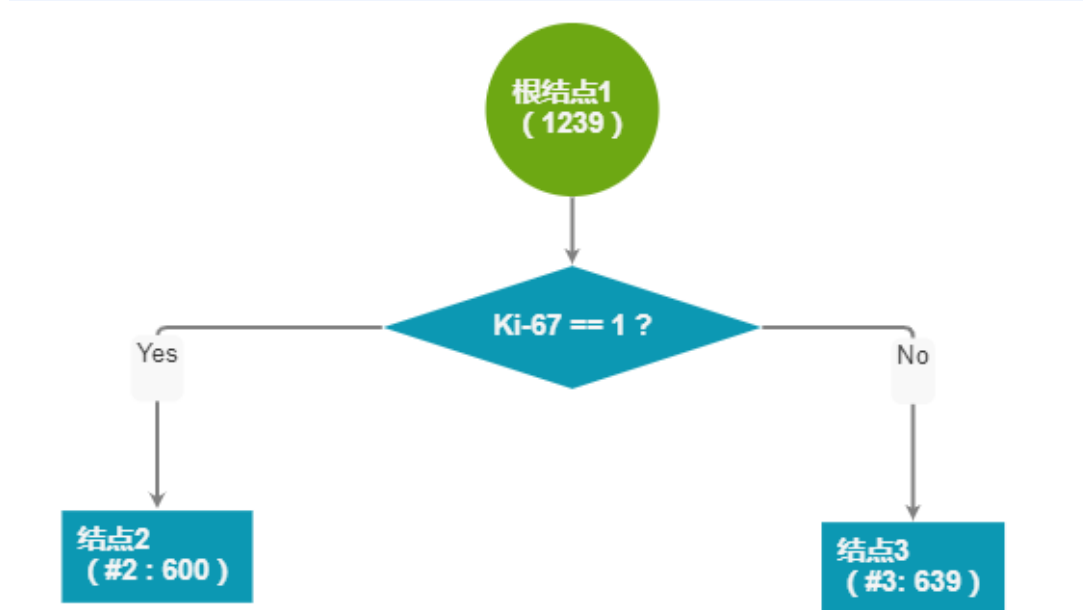


图 3-1 示例

为了找出使预后因素对生存影响差异最大的患者亚群，我们改进了 Breiman et al. 和 Zhang and Singer 提出的迭代划分技术。迭代划分的一个主要的动机是建立一个具有针对性的客观的系统来诊断特定的疾病和治疗手段^[12]。在这里，我们简单描述一下这一方法。

正如图 3-1 所展示的，迭代划分的目的是建立一个分类树来定义使预后因素对生存影响存在明显差异的亚群。我们从全部样本开始步骤，如图中结点 1 所示，这也通常被记为根结点。在我们的数据中，根结点由 1239 例患者样本组成，这个结点会分裂成两个子节点，标记为结点 2 和结点 3。这一分裂表示根据一个特定的评判准则，结点 2 和结点 3 比结点 1 表现更好。这里我们的评判准则为结点内所有样本建立的 COX 风险比例模型的似然函数值，值越大认为表现越好。我们考虑了数据中所有可能的预后因素以及每个因素可能的分裂方式，不同于像逻辑回归含参数的统计方法，迭代划分没有规定响应变量与协变量之间存在已知的关系形式。我们用每一个预后因素来给出一个

是或否的判定。举个例子，考虑年龄预后因素，我们可能会问是否患者的年龄是否大于 30 岁，若答案为是，则将该患者划分到结点 2，否则划分到结点 3。至于为什么我们要使用 30 岁作为划分点，则需要看 30 岁和 29 岁，28 岁，或其他患者中存在的年龄来根据评判准则进行比较，看是否 30 岁作为划分点时的表现最好。除此之外，我们是否选择年龄预后因素来划分根结点中的患者样本是基于它能不能创造一个表现更好的划分，相比于 Ki-67 蛋白质，乳腺癌肿瘤类型 tumor.grade 等。

如图 3-1，“Ki-67 蛋白质”为根结点表现最好的分裂变量，接下来我们重复划分根结点时的步骤继续寻找结点 2 和结点 3。迭代地划分，一颗初始的分类树就构建出来了。接着我们需要进行剪枝的步骤^[13]来去除多余的子结点。在本研究中，由于我们还需要保留预后因素变量来使用 COX 比例风险回归模型来衡量比较年龄预后因素对各患者亚群的生存影响大小，因此我们的剪枝原则为最多用三个预后因素变量来进行亚群的划分。

另外，为了避免患者亚群中的样本数过少导致没用临床意义，我们规定树结构中的每个结点的都需要有 50 个样本以上且树的最大层数为 3；若没有可能的分裂，则停止切分。

当找到特定的亚群之后，我们需要探究各个亚群下年龄预后因素对患者生存影响的差异。在这里，我们应用的是针对生存函数的 COX 比例风险回归模型来衡量各个预后因素对患者生存的影响。COX 模型的系数可以衡量一个协变量对整个模型生存风险的影响。举个例子，假设有两个患者，除了年龄之外所有的预后因素都相同，一个 30 岁，另一个 40 岁，我们将他们分别代入 COX 模型中，则它们的风险函数的比值可以由年龄预后因素的系数来表示出：

$$\frac{h(t, age = 40, Her2, Ki67, \dots, ER)}{h(t, age = 30, Her2, Ki67, \dots, ER)} = e^{10 * \beta_{age}}$$

另外，为了衡量找到特定亚群来分别建模的准确性，我们使用生存数据中经常使用的一致性指数 c-index 来比较分别建模与直接对全体患者样本建模的模型预测精度。

3.2 统计学知识点描述

3.2.1 COX 模型

在分析生存数据当中，风险比例模型是被应用的最广泛的步骤。若满足 PH(proportional hazards) 假设，则我们可以不考虑风险函数来求得估计参数，用于生存数据的这种求参

数的方法则称为 COX 比例风险模型。

设样本 i 的各个协变量的值为 $X_i = (X_{i1}, \dots, X_{ip})$, 则 COX 比例风险回归模型的风险函数在时间 t 对于患者 i 有以下形式:

$$h(t|X_i) = h_0(t) \exp(\beta_1 X_{i1} + \dots + \beta_p X_{ip}) = h_0(t) \exp(X_i * \beta)$$

其中: $h_0(\cdot)$ 是基准风险函数 (baseline hazard function);

患者 i 的事件被观察到在时间 Y_i 发生的概率可以被写成下列形式:

$$L_i(\beta) = \frac{h(Y_i | X_i)}{\sum_{j: Y_j \geq Y_i} h_0(Y_i | X_j)} = \frac{\theta_i}{\sum_{j: Y_j \neq Y_i} \theta_j}$$

其中, $\theta_j = \exp(X_j * \beta)$

这也称作偏似然函数, 可以用来求预后因素协变量的系数:

把所有患者样本看作相互独立的, 则所有事件发生的联合概率为:

$$L(\beta) = \prod_{i: M_i=1} L_i(\beta)$$

其中, 事件的发生用 $M_i = 1$ 来表示。

则对应的对数偏似然函数为:

$$l(\beta) = \sum_{i: M_i=1} (X_i * \beta - \log \sum_{j: Y_j \neq Y_i} \theta_j)$$

接着就可以使用 Newton-Raphson 法来求解预后因素协变量的系数^[14]。

3.2.2 决策树

决策树是一个递归地将自变量划分均匀区域的决策规则组成的分层模型^[15]。它是一个稳健的分类算法, 随着数据挖掘的发展变得越来越受欢迎。以决策树为基础的分类算法都有一个由结点 (或叶子) 和分支组成的树结构。这个树结构是根据一系列决策规则以特定的顺序来应用构建的; 分支 (即分裂) 是在树的某一特定的层上由不同的决策规则组成^[16]; 决策树的目的是寻找一些决策规则来预测输入变量的结果^[15]。如果从树结构中预测不同的场景, 则决策树可以解释为规则归纳技术。一个决策树由三种结点构成: 决策结点, 机会结点和结束结点。构造决策树的算法有分类树和回归树 (CART),

ID3s, CHAIDs, C4.5 和 C5.0 决策树^[17]。这些算法的目标是最小化树的大小,从而保证最终决策的质量,缩短决策过程。

本研究当中我们应用决策树的目的是寻特定的亚群使年龄对每个亚群中的患者生存影响差异最大,因此我们设定切分准则为最大化以下函数:

$$\max\left(\frac{\text{diff}_L * n_L + \text{diff}_R * n_R}{n_L + n_R}\right)$$

其中:

$$\text{diff} = L_p(\beta) = \prod_{j=1}^k \frac{\exp(z_j^T \beta)}{\sum_{t \in R(t_j)} \exp(z_t^T \beta)}$$

$\text{diff}_{L(R)}$: 表示切分后左(右)孩子结点的 diff 值;

$n_{L(R)}$: 表示切分后左(右)孩子结点的患者样本量;

$L_p(\beta)$: 表示 COX 比例风险回归模型的偏似然函数;

$z_j = (z_1, \dots, z_p)$: 表示第 j 个患者;

$\beta = (\beta_1, \beta_2, \dots, \beta_p)$: 表示 p 个协变量的系数值;

在本研究中我们考虑的预后因素作为协变量的有:

age; ER; PR; HER2; Ki67; surgery.type; histological.type; 共 7 个变量

对此构建的 COX 比例风险回归模型如下:

$$H(t|z_i) = h_0(t) * \exp(\text{age} * z_{i1} + \text{ER} * z_{i2} + \text{PR} * z_{i3} + \dots + \text{histological.type} * z_{i7}) = h_0(t) * \exp(z_i^T * \beta)$$

迭代地使用上述切分准则把初始的树结构构建好之后,接下来就需要进入剪枝步骤^[18]来去除多余的子结点,最后才能得到最终的决策树找出特定的患者亚群。在本研究中,由于我们还需要保留预后因素变量来使用 COX 比例风险回归模型来衡量比较年龄预后因素对各患者亚群的生存影响大小,因此我们的剪枝原则为最多用三个预后因素变量来进行亚群的划分。

另外,为了避免患者亚群中的样本数过少导致没用临床意义,我们规定树结构中的每个结点的都需要有 50 个样本以上且树的最大层数为 3;若没有可能的分裂,则停止切分。

3.2.3 一致性指数 (C-index)

一个具有很强区分能力的模型有望预测实际生存时间更长的患者生存时间更长。Harrell 的 C-index 是被应用得最广泛的测量生存模型预测准确率的指标。它是对任意两个患者 i 和 j 估计一致性概率 $P_r(W_i < W_j | Z_i < Z_j)$, 其中 W 是根据模型预测的生存时间, Z

是患者实际潜在的生存时间。在实际应用中, 预测患者个体的生存时间是一件困难的事情。Harrel(1996)^[19] 指出预测的生存时间可以和风险比例模型中预测的生存的概率互换。因此, 一致性概率, 记为 CP, 对于 $i \neq j$, 可以被写成以下形式:

$$\begin{aligned} CP &= Pr(W_i < W_j \mid Z_i < Z_j) = Pr(S_0(u)^{\exp(y_i * \beta)} < S_0(u)^{\exp(y_j * \beta)} \mid Z_i < Z_j) \\ &= Pr(y_i * \beta > y_j * \beta \mid Z_i < Z_j) \end{aligned}$$

Harrel 的 C-index 是在不考虑假定生存数据服从任何分布的前提下, 一致性概率的一个非参数估计。但是它有可能会被生存数据中相同对的数量。Chambless et al.(2011) 基于一个固定的生存函数推导出了另外一种估计方法来考虑删失事件存在的影响。若直接把有删失存在的患者对直接当作无用对来处理则会丢失很多的信息。因此, 我们修改 Chambless et al.(2011) 的方法, 通过假定生存指标 y_i 服从特定的先验分布来把删失的样本也考虑在内。我们假定生存数据服从分布 $f_\theta(y)$, 则我们可以把一致性概率的公式改写为:

$$CP = \frac{Pr(Y_i > Y_j, Z_i < Z_j)}{Pr(Z_i < Z_j)} = 2 \int_{-\infty}^{\infty} \int_{-\infty}^{y_i} Pr(Z_i < Z_j \mid Y_i = y_i, Y_j = y_j) f_\theta(y_i) f_\theta(y_j) dy_i dy_j$$

在比例风险模型下, Gonen 和 Heller(2005) 证明了

$$Pr(Z_i < Z_j \mid Y_i = y_i, Y_j = y_j) = \frac{1}{1 + \exp(\beta * y_j - \beta * y_i)}$$

因此我们可以进一步简化一致性概率如下:

$$2 \int_{-\infty}^{\infty} \int_{-\infty}^{y_i} \frac{f_\theta(y_i) f_\theta(y_j)}{1 + \exp(\beta * y_j - \beta * y_i)} dy_i dy_j$$

这个结果表明了一致性概率由生存数据的先验分布和表示生存数据和生存时间的联系的回归模型参数二者共同决定的。因此我们可以通过把最大似然估计 (MLE) 的参数 $\hat{\beta}$ 代入上式来求出 C-index。

这也暗示了若参数估计 $\hat{\beta}$ 的偏差 (bias) 可以忽略, 一致性概率可以被正确的估计即使存在删失的协变量和删失的生存指标结果。

第四章 实验与结果

4.1 结果呈现

根据决策树构建的结果，我们将患者分成了 4 个亚群，用来分类的预后因素有 Ki67 和年龄 age 和 Her2;

其中指定:

亚群 1 为预后因素 Her2 指标为 1，年龄小于等于 40 岁的患者;

亚群 2 为预后因素 Her2 指标为 0，年龄小于等于 40 岁的患者;

亚群 3 为 Ki67 指标为 1，年龄大于 40 岁的患者;

亚群 4 为 Ki67 指标为 0，年龄大于 40 岁的患者。

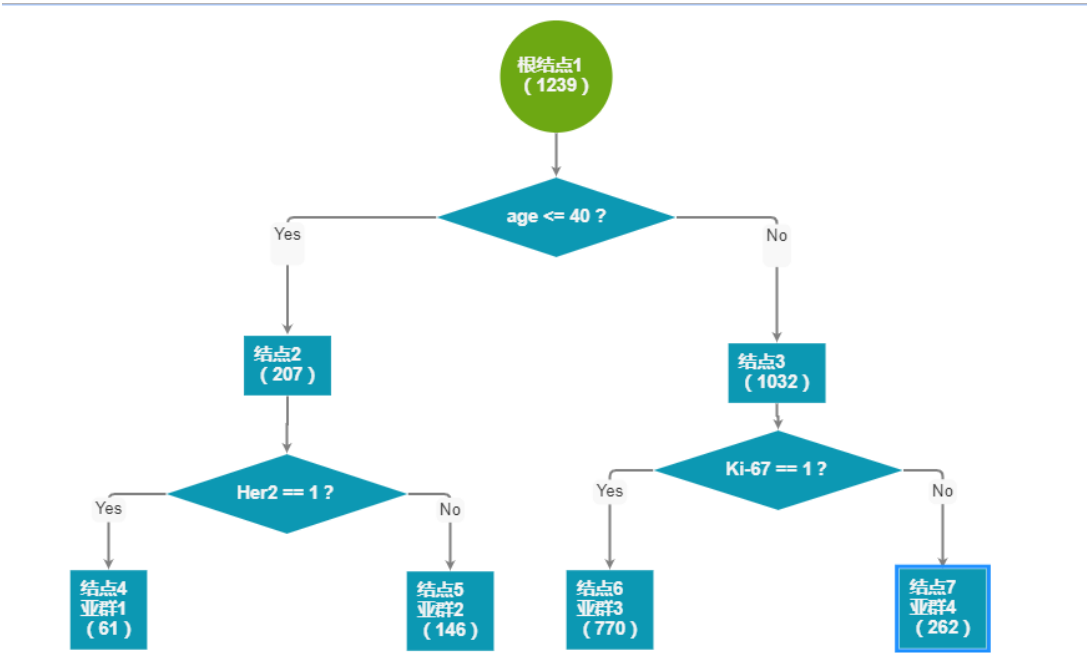


图 4-1 用于划分亚群的最终决策树

	type1	type2	type3	type4
numbers	61	146	770	262

每个亚群的患者样本数分别为 61，146，770，262;

4.1.1 生存分析

我们使用 Kaplan-Meier 方法来比较各个亚群之间的生存曲线图，结果呈现如下图，可以看到它们之间的不同，其中亚群 1 的患者的生存概率明显比其余亚群要低；亚群 4 的患者生存概率比其余亚群都要高。我们还对它们进行假设检验 (Log-Rank test) 零假设是组间不存在差异。

我们不仅可以从图中可以看出亚群组间生存曲线是有差异的，也从 Log-Rank test 的 P 值 0.0045 也可得出亚群组间有差异。



图 4-2 Survival Curve

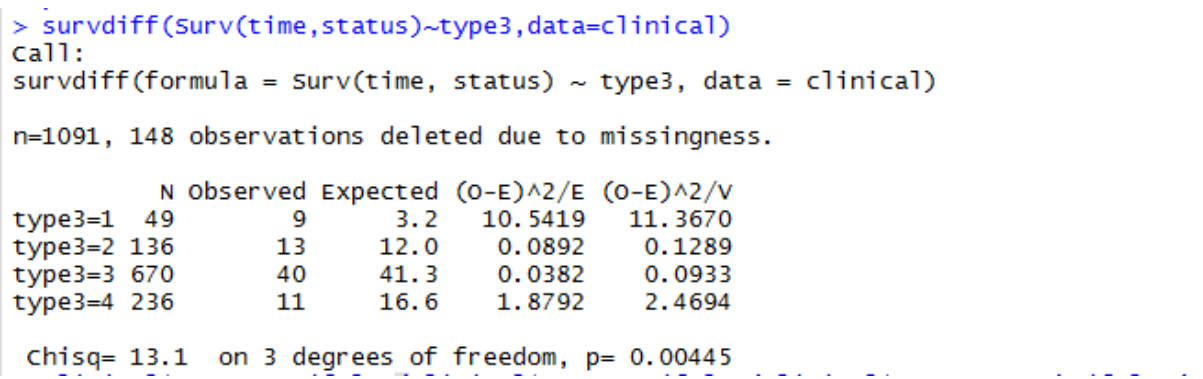


图 4-3 Summary of log-rank test

4.1.2 COX 模型参数分析

为了探究决策树找出的各个亚群之间年龄预后因素对生存的影响的差异大小，对亚群 1 和 2 其中的患者样本建立 COX 比例风险模型如下：

$$H(t) = h_0(t) * \exp(\beta_1 * age + \beta_2 * ER + \beta_3 * PR + \beta_4 * Ki67 + \beta_5 * histological.type + \beta_6 * surgery.type)$$

对亚群 3 和 4 其中的患者样本建立 COX 比例风险模型如下：

$$H(t) = h_0(t) * \exp(\beta_1 * age + \beta_2 * ER + \beta_3 * PR + \beta_4 * Her2 + \beta_5 * histological.type + \beta_6 * surgery.type)$$

在本研究中，我们以构建的 COX 比例风险模型中年龄预后因素协变量的系数来衡量年龄对各亚群生存的影响以及它们之间的差异 (coef of age)，具体的模型详细参数，每个系数的置信区间，显著程度等都在附录中展示；除此之外，我们还运用一致性指数 (c-index) 来衡量我们所建立的 COX 比例风险回归模型的预测精度：

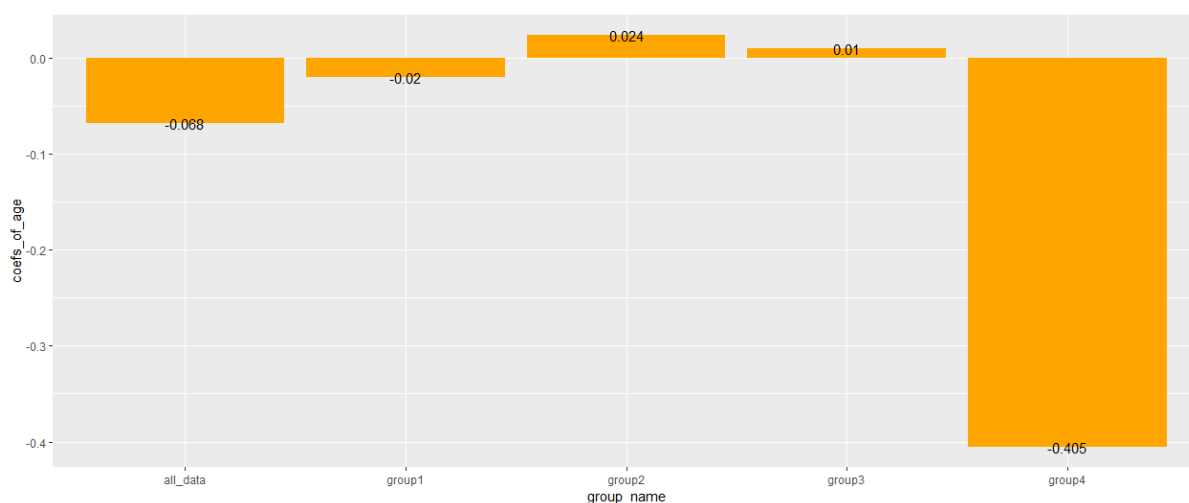


图 4-4 患者全体及各亚群年龄系数

	coef of age	c_index	se(c_index)
group1	0.024	0.730	0.13
group2	0.01	0.737	0.11
group3	-0.405	0.971	0.05
group4	-0.068	0.672	0.11
all_data	-0.02	0.654	0.04

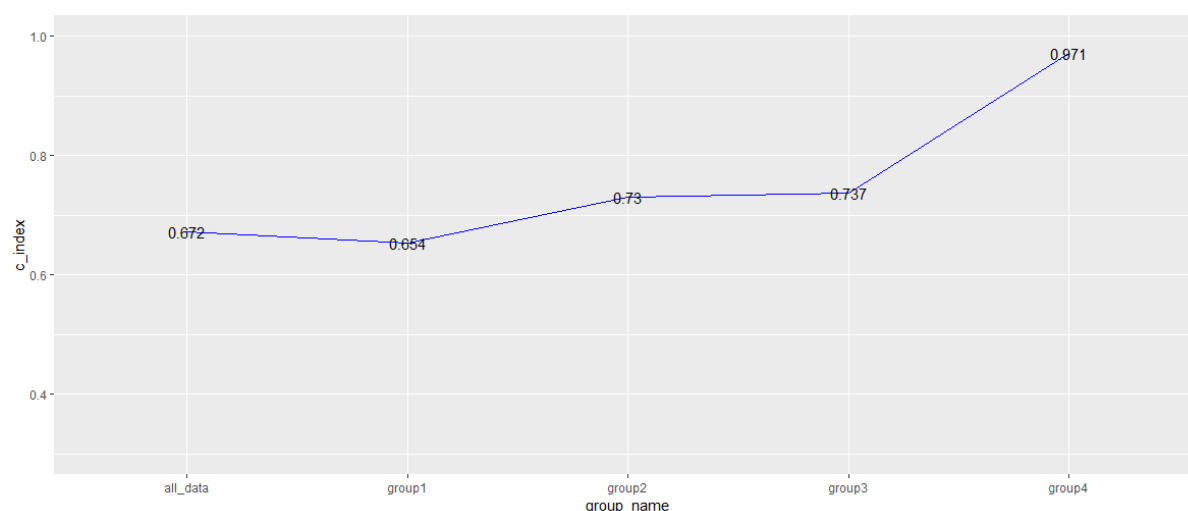


图 4-5 患者全体及各亚群模型 c-index 值

4.2 结果分析

由上面的表格可以看出：

(1) 对比于直接对所有的患者样本建立 COX 比例风险回归模型，可以看出切分成指定的 4 个亚群后，年龄预后因素对乳腺癌患者生存的影响明显存在差异。在亚群 1 和 2 中，模型内年龄预后因素的系数为正数（分别为 0.024 和 0.001），即说明在小于 40 岁以下的青年组乳腺癌患者中年龄会增加预后复发生存的风险，年龄越大，对患者的生存风险影响越大；并且，Her2 的表达状态也影响着小于 40 岁的青年组乳腺癌患者的年龄因素对生存风险的增加效果。根据我们的 COX 模型年龄系数的结果呈现，Her2 阳性的青年组患者比 Her2 阴性的患者的年龄因素对生存风险的促进作用影响更大。

(2) 同样的，在亚群 3 和 4 中，模型内年龄预后因素的系数为负数（分别为 -0.405 和 -0.068），即说明在 40 岁以上的中老年组乳腺癌患者中年龄会抑制预后复发生存的风险，年龄越大，对患者的生存风险影响越小；并且，Ki-67 蛋白质的表达状态也影响着 40 岁以上的中老年组乳腺癌患者的年龄因素对生存风险的抑制效果。根据我们的 COX 模型年龄系数的结果呈现，Ki-67 高表达（即指标为 1）的中老年组患者比 Ki-67 低表达（指标为 0）的患者的年龄因素对生存风险的抑制作用影响更大。

(3) 对比于直接对所有的患者样本建立 COX 比例风险回归模型，直接建模和亚群 1-4 的 c-index 分别为：0.654，0.672，0.737，0.971，0.730；因此将乳腺癌患者划分成这 4 个特定亚群后再建立的模型比直接建模的模型预测精度更高，对临床应用更加有意义。

4.3 结果讨论

我们在本研究中提出且用来构建分类树来找出特定患者亚群并建模衡量各自年龄对生存影响差异的方法是崭新的和数据驱动的。在我们的研究中，我们使用这 1239 例乳腺癌患者的数据展示验证了^[8]中报告的年龄对乳腺癌患者影响的关系。虽然我们的方法考虑了患者总体和亚群的分析，但它的合理性还需要更加深刻的去关注验证，以树为基础的方法在别的生物统计方面的研究已经很好地在 Zhang and Singer^[12] 的研究中建立了，并且我们的研究结果表明了该方法能够很好地为临床咨询和决策判断提供有用的信息。另外，我们所找到的特定的患者亚群并不是事先定义的，因此这些考虑让我们的方法十分地适合临床问题的解决，如针对某个特定的病人，将其归入我们的亚群分类中，然后选择合适的治疗方法。所以，本研究提出了一种重要的概念来验证临床实验中数据的有效性以及一种全新的方法来达到该目的。

在我们的决策树中使用的一些预后因素已经被以前的分析所证实过^[8]。我们了解到 Ki67 蛋白质表达的形态，Her2 表皮生长因子受体阴阳性，乳腺癌肿瘤的类型，乳腺癌肿瘤的大小等对乳腺癌患者的生存复发率存在影响，但是我们的研究中并不同于以往类似的研究^[8]，我们使用了 Her2 表皮生长因子受体阴阳性，Ki67 蛋白质表达的形态，age 三个预后治疗因素作为重要的划分变量来划分患者到特定的亚群，然后再使用剩下的预后治疗因素加上年龄建立 COX 风险比例模型。

其他的临床实验也曾证实了类似的预后治疗因素对乳腺癌患者的生存有影响，王仙玲等人^[20]的研究结果显示了肿瘤大小，Ki-67 和预后复发生存时间具有相关性 ($P < 0.05$)；在多因素分析中，肿瘤大小，Ki-67 均是 DFS 的独立预后因素 ($P < 0.05$)，提示原发肿瘤为 T1，淋巴结为转移以及 Ki-67 低表达患者预后较好；原发肿瘤体积较大，Ki-67 高表达患者预后较差，更易发生复发和转移。

王周权等人^[21]运用了 140 例老年女性乳腺癌患者作为研究对象，也证实了肿瘤大小，淋巴状态是独立的预后指标，以及由于老年患者预期寿命及高血压等合并症的存在，对于化疗这一治疗手段有一定的影响，化疗的药物的选择存在限制，甚至因此部分患者放弃了这一治疗手段，进而对其预后的复发生存时间存在一定的影响。因此，在我们的研究中，也将治疗手段当作其中一种预后因素。

另外，Jipeng et al.^[22] 也分析了乳腺癌患者根据免疫组织化学检测的雌激素受体 (ER)，人表皮生长因子受体 2 (HER2)，孕激素受体 (PR) 的表达情况进行分类管理，它的研究表明 TNBC 阳性（指 ER，PR，HER2 皆为阴性）的乳腺癌和 HER2 阳性的乳腺癌恶性程度比较高，容易复发，转移，预后较差，同时在病理学上都有肿瘤病灶高密度淋巴细胞浸润组织。因此本研究也同样把 ER，HER2，PR 这三者当作影响乳腺癌患者

预后生存复发的影响因素。

本研究仍旧存在一定的缺陷。首先所使用的方法和分析都是根据真实数据以及探索性的，必然存在固有的偏差；另外其他有用的信息，例如患者的淋巴管浸润和 p53 等信息都没有进行的分析；以及本研究使用的 1239 例患者数据收集是经过了约 15 年的时间跨度，在这段期间之中的治疗方案等的优化可能会对患者的预后产生影响。因此有必要对一个更大的患者样本及更长时间的记录的数据来进行类似相关的研究来验证我们的结论。

第五章 总结与展望

5.1 工作总结

乳腺癌仍然是目前医疗上十分关注的一种疾病，本文基于乳腺癌患者的复发生存数据，考虑了多种预后因素对患者生存的影响，以年龄预后因素的影响为切入点，创新地使用决策树方法根据预后因素 Ki67 与人表皮生长因子受体 Her2 以及年龄，将患者划分为 4 个特定的亚群；并且运用 COX 比例风险模型来探究在该 4 个特定亚群下年龄因素对患者生存影响的差异，然后使用一致性指数来比较划分亚群前后模型预测的精度高低。

总的来说，通过使用 1239 例乳腺癌患者生存数据进行实验，得出结论：Ki67 蛋白质高表达（指标为 1）及年龄大于 40 岁的患者亚群和 Ki67 蛋白质低表达（指标为 0）及年龄大于 40 岁的患者亚群中年龄预后因素对乳腺癌患者生存风险具有抑制、保护作用并且后者影响更大；Her2 阳性（指标为 1）及年龄小于等于 40 岁的患者亚群和 Her2 阴性（指标为 0）及年龄小于等于 40 岁的患者亚群中年龄预后因素对乳腺癌患者生存风险具有促进，增加作用并且后者影响更大；另外，根据模型的一致性指数，划分 4 个特定亚群后再建立的 COX 比例风险模型的预测精度会有明显提高，对临床应用更加有意义。

参考文献:

- [1] 刘秋华, 陈少芳, 杨芳, et al. 临床路径病友篇在手术治疗子宫肌瘤患者中的应用 [J]. 广东医学, 2016, **37**(z1):275–277.
- [2] Wang R, Lagakos S W, Ware J H, et al. Statistics in medicine—reporting of subgroup analyses in clinical trials[J]. New England Journal of Medicine, 2007, **357**(21):2189–2194.
- [3] DeSantis C, Ma J, Bryan L, et al. Breast cancer statistics, 2013[J]. CA: a cancer journal for clinicians, 2014, **64**(1):52–62.
- [4] Fredholm H, Eaker S, Frisell J, et al. Breast cancer in young women: poor survival despite intensive treatment[J]. PloS one, 2009, **4**(11):e7695.
- [5] Fernandopulle S M, Ang P C S, Tan P H. Breast carcinoma in women 35 years and younger: a pathological study[J]. Pathology, 2006, **38**(3):219–222.
- [6] Yancik R, Wesley M N, Ries L A, et al. Effect of age and comorbidity in postmenopausal breast cancer patients aged 55 years and older[J]. Jama, 2001, **285**(7):885–892.
- [7] 周春阳, 李伟华. 不同年龄乳腺癌患者病理特征及预后分析 [J]. 健康研究, 2018, **v.38; No.179**(2):60–62+66.
- [8] Zhong W, Tan L, You N, et al. Abstract P2-08-56: Effects of young age on prognosis in patients with node-negative tumors 2 cm or smaller breast cancer[J]. International Journal of Cancer, 2019.
- [9] Yancik R, Wesley M, Ries L, et al. Effect of age and comorbidity in postmenopausal breast cancer patients aged 55 years and older.[J]. JAMA, 2001, **285**(7):885.
- [10] Zhang H, Legro R S, Zhang J, et al. Decision trees for identifying predictors of treatment effectiveness in clinical trials and its application to ovulation in a study of women with polycystic ovary syndrome[J]. Human Reproduction, 2010, **25**(10):2612–2621.
- [11] Pesonen E, Eskelinen M, Juhola M. Treatment of missing data values in a neural network based decision support system for acute abdominal pain[J]. Artificial Intelligence in Medicine, 1998, **13**(3):139–146.
- [12] Zhang H, Singer B H. Recursive partitioning in the health sciences[M]. Springer Science & Business Media, 2013.
- [13] Legro R S, Barnhart H X, Schlaff W D, et al. Clomiphene, metformin, or both for infertility in the polycystic ovary syndrome[J]. New England Journal of Medicine, 2007, **356**(6):551–566.
- [14] Cox D R. Regression models and life-tables[J]. J. Roy. Statist. Soc. Ser. B, 1972, **34**:187–220.
- [15] Pradhan B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS[J]. Computers & Geosciences, 2013, **51**:350–365.
- [16] Petrović J, Ibrić S, Betz G, et al. Optimization of matrix tablets controlled drug release using Elman dynamic neural networks and decision trees[J]. International journal of pharmaceuticals, 2012, **428**(1-2):57–67.

- [17] Biggs D, De Ville B, Suen E. A method of choosing multiway partitions for classification and decision trees[J]. Journal of applied statistics, 1991, **18**(1):49–62.
- [18] Kaur A. Zhang H, Singer B, Recursive partitioning in the health sciences[J]. STATISTICAL METHODS IN MEDICAL RESEARCH, 2001, **10**:309–310.
- [19] Kim Y, Kong L. Estimation of C-index for cox proportional hazards model with censored biomarker covariate subject to limits of detection[J]. Journal of biopharmaceutical statistics, 2015, **25**(3):459–473.
- [20] 王仙玲. 三阴性乳腺癌的临床特征分析 [J]. 肿瘤研究与临床, 2015, **27**(3):201–203.
- [21] [1. 老年乳腺癌 140 例临床病理特点及预后分析 [J]. 陕西医学杂志, 2009, **38**(11):1504–1506.
- [22] 吉国, 龚悦, 狄根红. 肿瘤浸润淋巴细胞在三阴性乳腺癌和 HER2 阳性乳腺癌中作用的最新研究进展 [J]. 中国癌症杂志, 2018, **28**(12):928–934.

致谢

四年的本科生涯一眨眼就要过去了，回顾这段时间，我不仅在学习上学到了很多知识，培养了自己自我学习的能力，也更加了解了科研学术工作的意义和重要性。这一切离不开老师和同学们的悉心教导和帮助，在此衷心地想他们表达真诚的谢意。

首先要感谢我的指导老师尤娜教授。作为一名本科生，我的学术科研素养可能有所欠缺，缺少相关的研究经验，经常会遇到各种各样的问题，或者是抓不住重点，搞不清楚问题的本质重点是什么。但是尤老师对待我们十分热心，耐心地引导我们自己去发现问题并激发我们的灵感让我们自己去尝试解决，从而培养我们的科研能力。老师严谨的治学态度和耐心认真对待工作的方式是我学习的榜样，尤老师一直以来的督促和指导我会永远记在心里，在此向尤老师致以崇高的敬意和衷心的感谢。

另外我还要感谢我的好朋友们，不仅在学习上我们互相帮助，讨论解答疑惑；在颇有压力的学习环境下，是他们让我能放松自己，调整好心态，每天都积极面对科研和学习生活。

最后我还要感谢我的家人，正是家人们一直以来对我学业上的支持，支持我继续在学术上深造，我才有这么多好的机会来奋斗提升自己，才能获得目前的小小成果。感谢大家，也感谢四年来一直拼搏努力的自己。

陈敬贤

2019年5月1日

附录 A supplement

A.1 各个模型的具体参数展示

```
> summary(fitall)
Call:
coxph(formula = surv(time, status) ~ age + histological.type +
      ER + PR + Her2 + Ki67 + surgery.type, data = clinical)

n= 955, number of events= 62
(284 observations deleted due to missingness)

              coef exp(coef) se(coef)      z Pr(>|z|)
age          -0.02383   0.97645  0.01246 -1.912  0.0559 .
histological.type -0.34281   0.70977  0.22693 -1.511  0.1309
ER1          -0.17882   0.83625  0.37912 -0.472  0.6371
PR1          -0.39501   0.67367  0.35401 -1.116  0.2645
Her21         0.15716   1.17018  0.30009  0.524  0.6005
Ki671         0.38843   1.47467  0.34850  1.115  0.2650
surgery.type   0.14946   1.16121  0.09312  1.605  0.1085
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
age              0.9765      1.0241    0.9529    1.001
histological.type 0.7098      1.4089    0.4549    1.107
ER1              0.8363      1.1958    0.3978    1.758
PR1              0.6737      1.4844    0.3366    1.348
Her21            1.1702      0.8546    0.6499    2.107
Ki671            1.4747      0.6781    0.7448    2.920
surgery.type     1.1612      0.8612    0.9675    1.394

Concordance= 0.654 (se = 0.041 )
Rsquare= 0.019 (max possible= 0.548 )
Likelihood ratio test= 17.97 on 7 df,  p=0.01212
wald test               = 17.95 on 7 df,  p=0.01221
score (logrank) test = 18.17 on 7 df,  p=0.01122
```

图 A-1 基于全体样本建立的 COX 模型参数

```

              coef exp(coef) se(coef)      z Pr(>|z|)
age           0.02399   1.02428  0.04455  0.539   0.590
ER1          -1.19087   0.30396  1.10773 -1.075   0.282
PR1          -0.98472   0.37354  1.10136 -0.894   0.371
Her21         0.71981   2.05404  1.03110  0.698   0.485
histological.type 0.05277   1.05419  0.38905  0.136   0.892
surgery.type   0.13498   1.14452  0.27504  0.491   0.624

              exp(coef) exp(-coef) lower .95 upper .95
age              1.0243      0.9763   0.93865    1.118
ER1              0.3040      3.2899   0.03467    2.665
PR1              0.3735      2.6771   0.04314    3.235
Her21            2.0540      0.4868   0.27223   15.498
histological.type 1.0542      0.9486   0.49177    2.260
surgery.type     1.1445      0.8737   0.66759    1.962

Concordance= 0.73 (se = 0.108 )
Rsquare= 0.043 (max possible= 0.356 )
Likelihood ratio test= 7.46 on 6 df, p=0.2808
Wald test = 10.29 on 6 df, p=0.1128
Score (logrank) test = 14.72 on 6 df, p=0.02258

```

图 A-2 基于亚群 1 样本建立的 COX 模型参数

```

              coef exp(coef) se(coef)      z Pr(>|z|)
age           0.0005526   1.0005527  0.0567922  0.010   0.9922
ER1          -1.0901891   0.3361529  1.2008321 -0.908   0.3640
PR1          -0.9210441   0.3981032  0.9448748 -0.975   0.3297
Her21        -1.9475466   0.1426236  1.2526264 -1.555   0.1200
histological.type -0.7961761   0.4510504  0.3621371 -2.199   0.0279 *
surgery.type  -0.0838672   0.9195534  0.2514993 -0.333   0.7388
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
age              1.0006      0.9994   0.89516    1.1184
ER1              0.3362      2.9748   0.03194    3.5374
PR1              0.3981      2.5119   0.06248    2.5367
Her21            0.1426      7.0115   0.01224    1.6612
histological.type 0.4511      2.2170   0.22181    0.9172
surgery.type     0.9196      1.0875   0.56169    1.5054

Concordance= 0.737 (se = 0.1 )
Rsquare= 0.05 (max possible= 0.47 )
Likelihood ratio test= 7.59 on 6 df, p=0.2694
Wald test = 8.47 on 6 df, p=0.2054
Score (logrank) test = 8.14 on 6 df, p=0.2277

```

图 A-3 基于亚 2 样本建立的 COX 模型参数

```

              coef exp(coef) se(coef)      z Pr(>|z|)
age          -4.049e-01  6.671e-01  3.313e-01 -1.222  0.2216
ER1           9.998e-01  2.718e+00  2.116e+05  0.000  1.0000
PR1           1.992e+01  4.479e+08  1.000e+05  0.000  0.9998
Her21         1.402e+00  4.062e+00  1.885e+00  0.744  0.4571
histological.type -1.741e+01  2.758e-08  3.923e+04  0.000  0.9996
surgery.type   1.270e+00  3.561e+00  6.518e-01  1.949  0.0513 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
age          6.671e-01  1.499e+00    0.3485    1.277
ER1          2.718e+00  3.680e-01    0.0000    Inf
PR1          4.479e+08  2.233e-09    0.0000    Inf
Her21        4.062e+00  2.462e-01    0.1010   163.402
histological.type 2.758e-08  3.625e+07    0.0000    Inf
surgery.type   3.561e+00  2.808e-01    0.9926   12.778

Concordance= 0.971 (se = 0.21 )
Rsquare= 0.165 (max possible= 0.246 )
Likelihood ratio test= 12.29 on 6 df,  p=0.05586
wald test               = 4.47 on 6 df,  p=0.6129
Score (logrank) test = 17.22 on 6 df,  p=0.008507

```

图 A-4 基于亚群 3 样本建立的 COX 模型参数

```

              coef exp(coef) se(coef)      z Pr(>|z|)
age          -0.0676    0.9346  0.0197 -3.432 0.000599 ***
ER1          -0.1553    0.8562  0.4265 -0.364 0.715758
PR1          -0.3088    0.7343  0.4006 -0.771 0.440720
Her21         0.3126    1.3670  0.3410  0.917 0.359324
histological.type -0.1698  0.8439  0.2986 -0.569 0.569647
surgery.type   0.1207    1.1282  0.1205  1.001 0.316655
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
age          0.9346    1.0699    0.8992    0.9714
ER1          0.8562    1.1680    0.3711    1.9751
PR1          0.7343    1.3618    0.3349    1.6101
Her21        1.3670    0.7315    0.7006    2.6672
histological.type 0.8439    1.1850    0.4700    1.5150
surgery.type   1.1283    0.8863    0.8909    1.4288

Concordance= 0.672 (se = 0.051 )
Rsquare= 0.028 (max possible= 0.542 )
Likelihood ratio test= 16.35 on 6 df,  p=0.01199
wald test               = 16.58 on 6 df,  p=0.01097
Score (logrank) test = 17.3 on 6 df,  p=0.00824

```

图 A-5 基于亚群 4 样本建立的 COX 模型参数

毕业论文 (设计) 成绩评定记录
Grading Sheet of the Graduation Thesis (Design)

<p>指导教师评语 Comments of Supervisor:</p> <p>成绩评定 Grade:</p> <p>指导教师签名 Supervisor Signature: _____ Date: _____</p>					
<p>答辩小组或专业负责人意见 Comments of the Defense Committee:</p> <p>成绩评定 Grade:</p> <p>签名: _____ Date: _____ Signatures of Committee Members</p>					
<p>院系负责人意见 Comments of the Academic Chief of School:</p> <p>成绩评定 Grade:</p> <table style="width: 100%; border: none;"><tr><td style="width: 33%; padding-bottom: 10px;">签名 Signature: _____</td><td style="width: 33%; padding-bottom: 10px;">院系盖章 Stamp: _____</td><td style="width: 33%; padding-bottom: 10px;">Date: _____</td></tr></table>			签名 Signature: _____	院系盖章 Stamp: _____	Date: _____
签名 Signature: _____	院系盖章 Stamp: _____	Date: _____			