


 <https://gitee.com/lianxh/ST> 连享会 · 课程



# 事件史/生存分析

## Survival Analysis

2021 年  
4月24-25日



 [最新版课纲 PDF](#) | [课程主页](#)

## 目录

- 1. 课程概览
  - 1.1 课程提要
  - 1.2 报名
  - 1.3 听课软件
  - 特别提示
- 2. 嘉宾简介
- 3. 课程详情
  - 3.1 课程简介
  - 3.2 授课内容
  - 3.3 课程特色
  - 3.4 课前准备
    - 预备知识
    - 参考教材
- 4. 报名和缴费信息
  - 报名链接
  - 缴费方式
- 5. 诚聘英才 
- 相关课程

# 1. 课程概览

---

## 1.1 课程提要

- **课程主页:** <https://gitee.com/lianxh/ST>
- **简介:** 本课程介绍实证研究中常见的生存分析方法和模型, 包括: 描述性统计 (K-M 估计)、离散时间 Logit 模型、参数模型、分段恒定对数比率模型、比率风险模型等。
- **授课方式:** 网络直播+回放
- **时间:** 2021 年 4 月 24-25 日 (周六-周日)
  - **时段:** 上午 9:00-12:00; 下午 2:30-5:30; 课后 30-60 分钟答疑
- **授课嘉宾:** [王存同教授](#) (中央财经大学)
- **软件/课件:** Stata, 提供全套 Stata 实操程序、数据和 dofiles (开课前一周发送)。建议使用 Stata 15.0 或更高版本。

## 1.2 报名

- **报名优惠:** 两天回放 + 一个连享会[小直播课程](#)
- **报名链接:** <http://junquan18903405450.mikecrm.com/LyTjYYB>

## 1.3 听课软件

本次课程可以在手机, ipad 以及 windows 系统的电脑上听课。

- **手机/iPad:** 直接在应用商店搜索「**大黄蜂云课堂**」, 安装即可
- **电脑 (windows) :** <https://www.360dhf.cn/dhfplayer/>, 将本网址在浏览器打开, 下载「**大班课直播-学生端**」。

### 特别提示

一个账号绑定一个设备, 且听课电脑 **需要 windows 系统**, 请大家提前安排自己的听课设备。

- 一个网红能红多久？
- 婚姻难逃「七年之痒」吗？
- 「老字号」企业能撑多久？
- 一个上市公司退市的风险有多大？
- 一个刚确诊的癌症患者还有多少时日？

这些看似无关的现象背后，其实都涉及一个共同的核心问题：事件是否发生及何时发生或事件发生的可能性有多大。在多数研究中，我们仅关心某个时点或某个状态，却很少关心「持续时间」，而这恰恰是「生存分析」的优势所在。

当我们讨论一个上市公司是否退市时，是不包含时间特征的。但若考虑它何时退市或退市的风险有多大时，就要增加对时间维度的考量。换言之，我们要探索的因变量为「事件是否发生」以及「一个状态转换到另外一个状态间的持续时间」，并要考察是哪些因素影响了事件的发生速度及生存时间的长短，则与之对应的分析方法为「生存分析」(Survival Analysis)。

目前，「生存分析」受到了经济学、金融学、社会学、管理学等领域的重视和青睐。为此，我们邀请了中央财经大学的王存同教授与大家分享生存分析的基本概念、主要模型和 Stata 实现方法。

## 2. 嘉宾简介

**王存同**，中央财经大学教授。博士毕业于北京大学(与 University of Michigan 合作培养)，博士后研究员就职于美国伊利诺伊大学 (University of Illinois at Urbana-Champaign)。主要从事社会统计及计量经济分析、人口健康学等领域的研究与教学。2013 年入选教育部新世纪优秀人才计划，兼任北京大学社会科学方法培训教授、美国 PAA、国际 IUSSP 会员及 IUSSP 社会科学定量方法培训专家组成员、美国伊利诺伊大学合作研究员及中国青少年性健康教育委员会副主任委员等。曾在《中国社会科学》，《社会学研究》，*Chinese Sociological Review*, *Schizophrenia Research* 等期刊发表论文近百篇，出版著作 6 部；主持国家社科基金项目 2 项、横向课题 12 项。

**小插曲：**王存同教授长期兼任北京大学等多个高校社会科学方法的培训教授，授课生动活泼，通俗易懂，拥有众多粉丝。每逢王老师在中央财经大学沙河校区(距离市区 1.5 小时车程)开讲时，小伙伴们便像候鸟般赶赴沙河。每次上课总是座无虚席，或席地而坐、或背墙而立.....若想在教室前排的过道台阶上抢个位置，那可是要早早出动的。

## 3. 课程详情

### 3.1 课程简介

生存分析在不同的学科里有不同的名称。例如，在经济学、社会学及管理学科中也被称为「**事件史分析**」(Event History Analysis) 或「**久期分析**」(Duration Analysis)。

那么，用于生存分析的数据有什么特点呢？

以癌症患者为例，从确诊到死亡的时间是一个典型的连续数据。若研究者探索哪些因素(如治疗方案、年龄、性别等)影响该连续变量(即生存时间)，根据「**因变量的测量层次决定回归模型**」的基本原则，似乎 OLS 就可以胜任，但事实上远没有那么简单。这是因为部分癌症患者的生存时间很短，部分患者生存时间很长(甚至超过研究者的观察年限)，而我们限于观测时限、研究经费等因素，不可能永远跟踪每个患者并知道他/她确切的生存时间。因此，通常我们会有一人为设定的观测区间(如 4 年)。那么，在这个观测区间内，观测结果就会出现如下两种情况：

- 其一，若该患者在 4 年的观测区间内去世，则可知道他/她确切的生存时间；
- 其二，若该患者并未在观测区间内去世，则无法知道他/她确切的生存时间，但我们至少可以知道他/她的生存时间一定大于 4 年。这就出现了我们熟悉的数据缺失中数据被删截的现象，即出现了「**右删截数据**」；若这种数据缺失或删截与其它因素相关，则为数据被截除，即出现了「**截除数据**」。

数据被删截，以及时变变量(time-dependent covariate)的存在让我们很难再利用常规模型来处理事件史数据。例如，在癌症患者的案例中，我们感兴趣的事件是那些个体死亡的风险有多大，目的是研究有哪些因素影响了死亡的发生可能性以及什么时候发生。其中，部分自变量并不随时间发生变化，如性别、种族，而部分变量则随时间发生变化，如药剂量、婚姻状况、是否喝酒与吸毒等。

若采用二分类因变量 Logit 模型(死亡与否)，就会浪费关于何时死亡的时间信息。显然，在治疗之后一个月内就死亡的个体比那些在 48 个月内并没有死亡的个体，在死亡的可能性上是完全不同的。再如，我们可以建立一个代表患者从发病到死亡间所经历的时间的因变量，然后对这个因变量进行 OLS 分析。但那些在 48 个月之内都没有死亡的患者将会被「删截」，并将从分析中被剔除，从而导致估计偏倚。

由此可知，生存分析所对应的因变量是一个包含删截或截除数据的事件时间变量，即因变量同时包含了持续时间与事件发生的特征，为一个包含时间的复合变量。因此，对因变量考察时，类似对离散因变量的考察形式一样表现为概率模型，即包含「发生与否、何时发生」共性的概率。它包含了多种模型，如指数模型(Exponential Model)、冈珀茨(Gompertz)模型、韦伯(Weibull)模型、离散时间模型(Discrete-time Models)以及考克斯(Cox)模型等。

「**生存分析**」在近期的 Top 期刊中得到了越来越广泛的关注和应用，涉及经济学、社会学、人口学、政治学、管理学、心理学、法律学等诸多学科。

本课程力图使学员在较短时间内了解常见生存分析模型的基本思想、原理、条件及适用范围，并以真实数据为演示案例，培训学员模型构建、软件应用及结果解读的能力，提高学员定量论文写作的水平。

在授课时间安排上，将利用少量时间进行回归知识的回顾，重点讲述生存分析的原理、模型设定、结果解读和 Stata 实操。例如，描述性统计(K-M 估计)、参数模型、半参数模型、离散时间风险模型、分段指数模型等。

### 3.2 授课内容

本课程主要涵盖如下六个小专题。在讲解每个模型时，都将结合完整的 Stata 实现文档与学员共同进行实例操作，并提供完整的 .dta, .do 文档等资料，以便大家演练和应用。

各讲主要内容及参考文献列举如下：

**温馨提示：**下文列举的参考文献可以在线浏览(手机浏览时需安装 PDF 浏览器)，亦可以通过百度云盘或曲奇云盘下载：

- 百度云盘: <https://pan.baidu.com/s/149eF4l6GczGsaFiVll875g>; 提取码: 8mth
- 曲奇云盘: <https://quqi.gblhgk.com/s/880197/9zy1lpVxVreoRC8S>

## T1. 生存分析概述

- 生存分析的起源和基本概念
- 生存分析的应用场景
- 分析策略

## T2. 描述性分析: K-M 估计 (Kaplan–Meier estimator)

- Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American statistical Association*, 83(402), 414-425. [.PDF-](#)
- Peterson Jr, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *Journal of the American Statistical Association*, 72(360a), 854-858. [.PDF-](#)

## T3. 离散时间 Logit 模型 (Discrete-Time Logit Models)

- Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological methodology*, 13, 61-98. [.PDF-](#)
- Yamaguchi, K. (1990). Logit and multinomial logit models for discrete-time event-history analysis: a causal analysis of interdependent discrete state processes. *Quality and Quantity*, 24(3), 323-341. [.PDF-](#)
- Muthén, B., & Masyn, K. (2005). Discrete-time survival mixture analysis. *Journal of Educational and Behavioral statistics*, 30(1), 27-58. [.PDF-](#)
- Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of educational statistics*, 18(2), 155-195. [.PDF-](#)

## T4. 参数模型 (Parametric Models)

主要介绍包含加速失效时间模型 (Accelerated Failure Time Models, AFT)。例如, 对数正态模型 (Log-normal Models)、指数模型 (Exponential Models)、伽马模型 (Gamma Models)、Gompertz 模型及 Weibull 模型等。

- Lambert, P. C., & Royston, P. (2009). Further development of flexible parametric models for survival analysis. *The Stata Journal*, 9(2), 265-290. [.PDF-](#)
- Nelson, C. P., Lambert, P. C., Squire, I. B., & Jones, D. R. (2007). Flexible parametric models for relative survival, with application in coronary heart disease. *Statistics in Medicine*, 26(30), 5486-5498. [.PDF-](#)

## T5. 分段恒定对数比率模型 (Log-Rate Models for Piecewise Constant Rates)

- Lalive, R., Van Ours, J., & Zweimüller, J. (2006). How changes in financial incentives affect the duration of unemployment. *The Review of Economic Studies*, 73(4), 1009-1038. [.PDF-](#)
- Walder, A. G., Li, B., & Treiman, D. J. (2000). Politics and life chances in a state socialist regime: Dual career paths into the urban Chinese elite, 1949 to 1996. *American Sociological Review*, 191-209. [.PDF-](#)

## T6. 比例风险模型 (Proportional Hazards Models)

主要包括 Cox 模型(Cox Model) 和分层 Cox 模型 (Stratified Cox Models)。

- Pourhoseingholi, M. A., Hajizadeh, E., Moghimi Dehkordi, B., Safaee, A., Abadi, A., & Zali, M. R. (2007). Comparing Cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pacific Journal of Cancer Prevention*, 8(3), 412. [.PDF-](#)
- Tian, L., Zucker, D., & Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American statistical Association*, 100(469), 172-183. [.PDF-](#)

### 3.3 课程特色

- **课程定位明确。**并非专业统计学课程，课堂上不会有烦琐的公式推导，重在应用，即在基本熟悉各模型思想的基础上，进行数据分析及模型解读，实现理论与应用相结合。
- **讲解重点突出。**讲解模型时，重点关注学术论文常用中高级模型，及其应用范围、条件、数据处理难点及模型参数的解读。
- **课件细致实用。**如各模型都附带完整的 Stata 实现过程 (do 文档) 和真实数据，学员无需进行烦琐编程，即可快速实现模型运用。
- **深度互动。**在课后答疑环节和课程微信群中，学员可就自己的研究数据进行沟通和解疑。
- **论文经验分享。**如与学员分享定量研究论文写作、投稿、修改，以及与编辑部、导师和合作者沟通的实用技巧。
- **讲师风格鲜明。**王老师风趣幽默，感染力强。在北京大学主讲全国高校教师社会科学定量方法暑期培训课时，多名学员评价其「统计课程如小说般引人入胜，不忍下课」、「讲解系统且实用性强」，已有多名学员应用所学成功发表量化研究论文。

### 3.4 课前准备

#### 预备知识

希望学员在开课前能重温计量经济学基础知识，对线性回归、假设检验有所了解，并具备一定的 Stata 软件操作基础。若对社会调查方法、社会统计有所了解，则更容易理解课程中的案例。当然，「大道至简、殊途同归」，

#### 参考教材

前 3 本为生存分析领域经典教科书，后 2 本偏向于 Stata 应用。

- Allison, P. D., 1984. *Event History Analysis* (2nd) (No. 46). Sage. [.在线阅读-](#)
- Yamaguchi, K., 1991. *Event History Analysis*. Sage. [.Link-](#), [.PDF-](#)
- Singer, J. D., Willett, J. B., & Willett, J. B., 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford university press. [.Link-](#)
- Cleves, Mario, William Gould, and Yulia Marchenko. *An Introduction to Survival Analysis Using Stata*, Revised Third Edition. 2016. Stata Press. [.Link-](#)
- 陈强. 2014. *高级计量经济学及 Stata 应用* (第 2 版). 高等教育出版社. [.Link-](#)

## 4. 报名和缴费信息

- **主办方：**太原君泉教育咨询有限公司
- **标准费用** (含报名费、材料费)：2200 元/人 (全价)
- **优惠方案：**
  - 三人及以上团购/连享会现场班以及线上直播大课学员：9 折
  - 充值连享会会员：85 折
  - **温馨提示：** 以上各项优惠不能叠加使用。
- **联系方式：**
  - 邮箱：[wjx004@sina.com](mailto:wjx004@sina.com)
  - 电话 (微信同号)：王老师 18903405450；李老师 18636102467

### 报名链接

报名链接：<http://junquan18903405450.mikecrm.com/LyTjYYB>  
或 长按/扫描二维码报名：



### 缴费方式

#### 方式 1：对公转账

- 户名：太原君泉教育咨询有限公司
- 账号：35117530000023891 (晋商银行股份有限公司太原南中环支行)
- **温馨提示：** 对公转账时，请务必提供「**汇款人姓名-单位**」信息，以便确认。

#### 方式 2：微信扫码支付



**温馨提示：** 微信转账时，请务必在「添加备注」栏填写「**汇款人姓名-单位**」信息。



## 5. 诚聘助教

### 说明和要求

- **名额:** 6名
- **任务:**
  - **A. 课前准备:** 协助完成 1 篇推文, 风格类似于 [lianxh.cn](http://lianxh.cn) ;
  - **B. 开课前答疑:** 协助学员安装课件和软件, 在微信群中回答一些常见问题;
  - **C. 上课期间答疑:** 针对前一天学习的内容, 在微信群中答疑 (8:00-9:00, 19:00-22:00); 参见 [往期答疑](#)。
  - Note: 下午 5:30-6:00 的课后答疑由主讲教师负责。
- **要求:** 热心、尽职, 熟悉 Stata 的基本语法和常用命令, 能对常见问题进行解答和记录。
- **特别说明:** 往期按期完成任务的助教自动获得本期助教资格, 不必填写申请资料, 直接联系连老师即可。
- **截止时间:** 2021 年 3 月 15 日 (将于 3 月 17 日公布遴选结果于连享会主页: [lianxh.cn](http://lianxh.cn))

申请链接: <https://www.wjx.top/jq/75688406.aspx>

或扫码填写助教申请资料:



课程主页: <https://gitee.com/lianxh/ST>



## 相关课程

连享会-直播课 上线了!

<http://lianxh.duanshu.com>

免费公开课:

- [直击面板数据模型](#) - 连玉君, 时长: 1小时40分钟, [课程主页](#)
- [Stata 33 讲](#) - 连玉君, 每讲 15 分钟.
- [Stata 小白的取经之路](#) - 龙志能, 时长: 2 小时, [课程主页](#)
- 部分直播课 [课程资料下载](#) (PPT, dofiles等)

 <https://gitee.com/lianxh/text>

# 文本分析 · 爬虫 · 机器学习

2021 年  
3月27日开课



Text is data



## 关于我们

- **Stata连享会** 由中山大学连玉君老师团队创办, 定期分享实证分析经验。 [直播间](#) 有很多视频课程, 可以随时观看。
- [连享会-主页](#) 和 [知乎专栏](#), 400+ 推文, 实证分析不再抓狂。
- **公众号关键词搜索/回复** 功能已经上线。大家可以在公众号左下角点击键盘图标, 输入简要关键词, 以便快速呈现历史推文, 获取工具软件和数据下载。常见关键词: [课程](#), [直播](#), [视频](#), [客服](#), [模型设定](#), [研究设计](#), [stata](#), [plus](#), [绘图](#), [编程](#), [面板](#), [论文重现](#), [可视化](#), [RDD](#), [DID](#), [PSM](#), [合成控制法](#) 等