

《国际关系定量分析基础》本科课程大纲

清华大学社会科学学院

2021 秋季学期

授课教师: 陈冲

教师电邮: chongchen@tsinghua.edu.cn

教师办公时间: 星期二上午 9:00-11:00 或邮件预约

教师办公室: 明斋 312

助教: 胡竞天 (hjt19@mails.tsinghua.edu.cn)

课程网页: learn.tsinghua.edu.cn

上课时间: 星期一第三大节 13:30-15:05

上课教室: 法图 B110

助教: 郝梦圆 (hmy21@mails.tsinghua.edu.cn)

课程简介

本课程是为国际关系及其他社会科学专业高年级本科生开设的专业基础课,旨在介绍定量社会科学的基本方法和运用,着重培养学生利用数据和统计编程语言解决政治学、国际关系等社会科学研究中的实际问题的能力,培养学生适应和满足当今“数据科学(data science)”这一行业对于社会科学专业学生的基本需求。由于这是定量研究方法的入门应用性课程,因此并不要求学生拥有较好统计学知识(虽然熟悉统计学知识会增加学生的学习效果)。这门课会通过介绍 **R** 语言,关注如何使用 **R** 实现数据的处理和分析,以及如何使用数据与统计方法解决国际关系相关的现实问题。课程的主要内容包括 **R** 语言在政治学和国际关系研究中的运用、数据清理、数据转换、数据可视化、线性与非线性回归分析模型的应用和解读。与此同时,作为配套工具,课程会介绍如何使用 RMarkdown、L^AT_EX 等工具。

课程目标

课程结束时,学生被期望能阅读和理解国际关系主流期刊上的定量研究论文,能运用所学方法去复制和延伸本领域的相关研究。同时,本课程也将为学生研修更高阶的社会科学方法奠定基础。课程将贯彻“边做边学”(learn by doing)这一理念,通过教师讲授、学生练习、学生合作、师生互动等方式,增加学习效果。课程分为两大部分:第一部分着重介绍如何使用 **R** 软件清理和管理数据;第二部分着重介绍线性回归和非线性回归模型的估计、诊断以及解释。整个课程将聚焦国际关系领域内实际案例的应用和讲解,特别是通过解构和还原顶级期刊上的经验研究论文,向学生展示如何开展高质量的定量研究。具体而言,在课程结束时,学生被期望达到以下目标:

- 能够掌握定量社会科学的基本概念、工具和研究路径
- 能够熟悉国际关系研究常用的数据来源、产生过程和主要优缺点
- 能够熟练地使用 **R** 软件,满足“数据科学家”(data scientist)这一职业的基本要求
- 能够阅读和批判一流国际关系期刊上的定量论文
- 能够独立收集、清理和分析数据,完成高质量的定量论文
- 具备较强的合作能力,适应团队工作,应用定量方法解决现实世界的问题

教材和阅读材料

本课程没有必需的指定教材,所有阅读的材料将会提前在教学系统中公布。鼓励学生在上课之前阅读该周指定的材料。本课程的部分阅读来自于以下推荐的教材:

- [新西兰] 哈德利·威克姆、[美] 加勒特·格罗勒芒：《R 数据科学》（陈光欣译），人民邮电出版社，2018 年。
- Nina Zumel and John Mount, *Practical Data Science With R* (Manning; Second Edition, 2020)
- Kosuke Imai, *Quantitative Social Science: An Introduction* (Princeton University Press, 2017)
 - 中译本：[日] 今井耕介：《量化社会科学导论》（祖梓文、徐轶青译），上海财经大学出版社，2020 年。
- Eric Goh Ming Hui, *Learn R for Applied Statistics With Data Visualizations, Regressions, and Statistics* (Apress, 2019)

关于线上教学

本课程将根据疫情防控情况，采取融合式线上线下教学。线上教学将配合使用荷塘雨课堂和腾讯会议。

- 腾讯会议
 - 加入链接：
 - 会议 ID：
 - 会议密码：
- 荷塘雨课堂
 - 班级号：

统计软件、数据和代码

本课程将使用免费的统计软件 **R** 和 **RStudio**。**R** 是目前一款非常主流的应用统计软件，具有强大的统计计算能力。课程初期将会介绍如何使用 RMarkdown，学生提交的作业和论文均需通过 **R** 和 **RStudio** 完成，利用 RMarkdown 提交，确保 100% 的可复制性。安装过程如下（根据电脑操作系统选择对应的版本）：

- 首先安装 **R**，下载网址 <https://www.r-project.org/>。
- 其次安装 **RStudio**，下载网址 <https://rstudio.com/>。

课程将使用国际关系和政治学研究中的真实数据。课程中的讲义 (slides)、**R** 代码和数据将会在课前公布在网络学堂或课程的网站和 GitHub，鼓励学生在课前课后使用讲义中的数据 and 代码练习使用。

课程考核与课堂政策

本课程以教师讲授 (lecture) 形式为主，学生需要积极参与到课堂讨论与展示中，并需要积极与小组成员合作。本课程没有考试，学生的最终成绩取决于个人部分（占 40%）和小组部分（占 60%）。学生最终成绩按照清华大学教务处相关规定转换成等级制。

成绩构成

- 个人部分（40%）
 - 课堂参与（10%）：出勤、发言和提问。如有正当原由确实无法到课，须提前以 Email 形式向教师书面请假，说明事由

- 个人课后作业 (30%): 学生有 2 次个人独立完成的课后作业, 每次作业占 15%
- 小组部分 (60%)
 - 小组期中考试 (20%): 在第 7 周结束之后学生以小组为单位, 完成一次以数据清理与可视化为主要内容的“回家”开卷 (take-home open book) 考试。
 - 小组期末论文 (30%): 期末考核为一篇 10-15 页的小组研究论文
 - 小组课堂展示 (10%): 学期末最后一次课为小组课堂展示, 学生以小组为单位报告小组期末论文, 每组时间约为 10 分钟
- 注意事项:
 - 分组: 根据选课人数, 学生会按照 3-5 人一组的的方式完成课后作业和期末论文。分组结果将在第 3 周 (9 月 27 日) 课上宣布, 鼓励学生自由组合。
 - 作业: 课程的课后作业以 R 实践为主, 强调学生对模型的理解和解读, 学生通常有两周的时间完成课后作业。课后作业和期末论文的截止日期为指定日期当日中午 (11:59am)。迟交或者不交将会按照每天扣除 10% 的比例扣除该部分的成绩, 迟交一星期该部分将不会有成绩。如有特殊情况, 需要在截止日前 24 小时以 email 形式正式获得教师许可。
 - 期中考试: 学生以小组为单位完成期中考试, 考试的形式为“回家”开卷 (take-home open book) 方式。学生将有 72 小时完成期中考试 (10 月 29 日中午 12 点-11 月 1 日中午 12 点), 考试内容为数据清理与可视化操作。
 - 期末论文: 学生以小组形式自由选定一个研究问题, 提出自己的研究假设, 并通过收集、清理、和探索性分析数据以及回归分析, 完成假设检验, 报告研究结果。论文正文长度在 10-15 页之间 (包括图表、参考文献在内)。论文正文一般包括研究问题、理论和研究假设、数据和方法、统计分析和讨论等部分。学生在第 6 周结束之后需要向教师报告论文选题, 获得许可后方可进行
 - 课程作业和期末论文均需使用 RMarkdown 完成, 教师会提供相应的模板, 同时鼓励小组作业和期末论文的合作通过 GitHub 平台完成。期末论文同时应提交相应的复制材料 (replication materials) 如数据、R 代码等。
 - 所有课程作业和期末论文均通过清华大学网络学堂提交

等级制成绩

课程最终成绩采取等级制, 教师在学期末根据学生的表现可能会进行相应调整 (grading on a curve), 每个等级人数没有限制。具体标准如下:

≥ 93	A	70 - 72	C
90 - 92	A-	67 - 69	C-
85 - 89	B+	63 - 66	D+
80 - 84	B	60 - 62	D
77 - 79	B-	≤ 59	F
73 - 76	C+		

手机和电脑使用规则

课堂上严禁使用手机开展与课程无关活动, 请在上课之前将手机调为静音。学生请尽量携带电脑或平板电脑到课堂。

Email 政策

Email 是本课程师生交流的主要方式。学生来信请以“IR300_ 姓名”方式作为邮件标题开头。我在工作日会及时回复学生来信。考虑到今年疫情防控的特殊性，课程会兼顾使用微信、网络学堂等方式进行课外互动。

课后如何寻求帮助

每周二下午 1:00-3:00 为教师开放交流时间 (office hours)，鼓励学生在此时间段到教师办公室进行交流。若以上时间存在冲突，学生也可以通过邮件与教师另行约定时间。同时鼓励学生自主在网络上寻求问题答案，大部分的 debugging 问题其实已经存在于网络上 (例如 <https://stackoverflow.com>)，如何将问题用准确的语言表述出来既是挑战，也是本课程的训练目标之一。例如：

- [Taking Good Notes](#)
- [Reading a Regression Table: A Guide for Students](#)

诚信、多样性与包容性政策

学生应严格遵守清华大学相关规定，在作业和期末论文中严禁抄袭、剽窃行为。课堂坚持包容和多样性原则，学生在小组活动中应坚持平等、互相尊重原则。

课程日历

鼓励学生在课前阅读每周指定的材料，在课后按时完成作业。课程的阅读材料有可能根据课程的进度做细微调整。鼓励学生及时查看网络学堂或者教师的邮件通知。所有阅读材料将提前在[网络学堂](#)发布。

第一部分：R 与数据清理

第 1 周, 2021-09-13: 国际关系定量分析导论

阅读材料:

- H. Akn Ünver. “Computational International Relations: What Can Programming, Coding and Internet Research Do for the Discipline?” *All Azimuth: A Journal of Foreign Policy and Peace* 8.2 (2019): 157-182.

第 2 周, 2021-09-20[中秋节调课，改为 9 月 18 日星期六]: R/RStudio & RMarkdown 操作

阅读材料:

- Zumel and Mount, chapter 2; chapter 11

第 3 周, 2021-09-27: 数据可视化 I

阅读材料:

- Eric Goh Ming Hui, chapter 5

第 4 周, 2021-10-04: 国庆假期 (无课)

第 5 周, 2021-10-11: 数据可视化 II

阅读材料:

- Eric Goh Ming Hui, chapter 5
 - 小组课后作业 (一)

第 6 周, 2021-10-18: 数据管理与清理 I

阅读材料:

- Zumel and Mount, chapters 4-5
– 期末论文选题

第 7 周, 2021-10-25: 数据管理与清理 II

阅读材料:

- Zumel and Mount, chapters 4-5
– 小组期中考试

第二部分：回归模型

第 8 周, 2021-11-01: 假设检验

阅读材料:

- Eric Goh Ming Hui, chapter 6
- Kosuke Imai, chapter 7.2

第 9 周, 2021-11-08: OLS 线性回归模型 I

阅读材料:

- Kosuke Imai, chapter 7.3
- Matthew Fuhrmann. “When Do Leaders Free-Ride? Business Experience and Contributions to Collective Defense.” *American Journal of Political Science* 64.2(2020): 416–431.

第 10 周, 2021-11-15: OLS 线性回归模型 II

阅读材料:

- Kosuke Imai, chapter 7.3

第 11 周, 2021-11-22: Logit & Probit 回归模型 I

阅读材料:

- Zumel and Mount, chapter 7.2
- J. Lawrence Broz, Zhiwen Zhang, and Gaoyang Wang. “Explaining Foreign Support for China’s Global Economic Leadership”. *International Organization* 74.3(2020): 417–52.

第 12 周, 2021-11-29: Logit & Probit 回归模型 II

阅读材料:

- Zumel and Mount, chapter 7.2
– 个人课后作业 (二)

第 13 周, 2021-12-06: 模型诊断与评估

阅读材料:

- Brian Greenhill, Michael D. Ward and Audrey Sacks. “The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Models.” *American Journal of Political Science* 55.4 (2011): 991-1002
- John Fox. *Regression Diagnostics: An Introduction* (second edition), Sage, 2020

第 14 周, 2021-12-13: 边际效应

阅读材料:

- Daniel Lüdtke. “Plotting Marginal Effects of Regression Models”
- Frederick Solt and Yue Hu. “dotwhisker: Dot-and-Whisker Plots of Regression Results.”
- Chong Chen. “**postregplots: An R Package for Visualizing Simulation-based Marginal Effects from Generalized Linear Models.**”

第 15 周, 2021-12-20: 面板数据分析

阅读材料:

- Christoph Hanck et al., “Chapter 10: Regression with Panel Data”, in *Introduction to Econometrics with R*, 2020.
- Kaitlyn Webster, Chong Chen, and Kyle Beardsley. “Conflict, Peace, and the Evolution of Women’s Empowerment.” *International Organization* 73.2(2019): 255-289.

第 16 周, 2021-12-27: 期末小组期末论文展示 (根据选课人数和小组数量, 期末展示时间可能进行相应的调整)

- **2022 年 1 月 3 日中午 11:59am : 小组期末论文截止**