

治理技术专题：政治数据分析

胡悦

2019-09

E-mail: yuehu@tsinghua.edu.cn

网站: learn.tsinghua.edu.cn

开放时间: 周一, 8:00–9:00 或预约

上课时间: 周一, 13:30–16:05

办公室: 明斋 114

教室: 四教 4202

课程概述

《政治数据分析》定位于“科学研究方法入门”，是政治学专业的必修课程，主要讲授与政治及广义社会科学研究相关的统计学知识以及基于 R 语言的相关应用。在内容层次上，包含对实证主义与人文主义、价值中立与价值关联、归纳逻辑与演绎逻辑、理论建构与理论检验的哲学思考；在方法层次上，涉及实现实验、调查、田野、文献等多种研究方法的基础逻辑和原则；在实践层次上，注重培养学生社会科学研究素养以及运用数据分析技术和分析方法解决实际问题的能力。

Analysis of political data is a gateway course of political methodology and compulsory for students in Political Science. The course introduces students to statistical and econometric knowledge for the analysis of quantitative data and provides a basis for more advanced statistical methods. After covering the fundamental concepts of statistics, such as probability and distributions, we will explore the classical normal linear regression model and its assumptions. We will also discuss the consequences and remedies for violations of these assumptions, including omitted variables, heteroscedasticity, autocorrelation, and endogeneity. If time permits, we will further talk about the use and interpretation of continuous, ordinal, nominal, and indicator variables, variable interactions, missing data and basic models for limited dependent variables. Along the way and primarily in the lab session, students will learn the basics of data collection, organization and management, measurement, data visualization and display, and univariate, bivariate and multivariate descriptive statistics in R.

以下内容将随课程进行随时调整！

课程目标

1. 了解基础概率论和计量经济学概念、原则和技法；

2. 准确理解 Ordinary Least Square (OLS) 假定；能使用 OLS 进行多元模型假设检验，并能加以正确解释；
3. 使用 R 进行简单的统计数据分析；
4. 提升规范学术写作水平。

参考书目

Moore, Will H., and David A. Siegel. 2013. *A Mathematics Course for Political and Social Research*. Princeton, NJ: Princeton University Pres.

Wonnacott, Thomas H., and Ronald J. Wonnacott. 1978. *Introductory Statistics*. Wiley.

Wooldridge, Jeffrey M. 2016. *Introductory Econometrics: A Modern Approach*. Sixth edition. Boston, MA: Cengage Learning.

陈正昌, and 贾俊平. 2016. 统计分析与 R. 中国人民大学出版社.

课程要求

本课原则有二：于我，勉力教学、力求公平；于诸位，各取所需，各尽所能。望诸生慎独，对任何形式作弊、抄袭零容忍。具体安排下详。

成绩分配

- **10%** 来自课堂参与。出勤为基本要求。每人有三次豁免机会，无需理由，无需告知。三次以外，每次缺勤减 1%，直至 10% 减完为止。请注意：三次豁免以外，除非学校通知停课，缺勤即减分，其中包含由于个人、家庭或其他缘由。请善用三次免责机会。
- **20%** 为 R 语言自学汇报表现。本课除安排必读材料外，还会根据当周内容安排相应的 R 语言操作内容，即“Bonus”部分，供学生在课后对该部分进行自修。自第四周起，本课将安排特定学生在每节课前对上一次课内容在 R 语言应用的自学情况进行 15–20 分钟的汇报。主讲学生应在当日课前将汇报材料交教师审定认可。教师会根据准备情况和汇报表现予以打分。
- **30%** 来自作业表现。本课依据课程进度会安排 3–5 次作业。每次作业均要求计算机缮打，并在规定日期当日课前邮件上交。邮件及作业附件均统一为“Method_ 姓名 _ 作业 *n*”。递交时间以教师邮箱接到邮件时间戳为准。一日以内晚交扣除当次作业总分 10%，一日至两日扣 20%，以此类推。
- **40%** 为期末研究项目表现。项目要求对一篇在政治学学术期刊上发表的使用线性回归的研究进行复刻研究 (reproducing)。需上交一篇 3000–5000 字的研究报告 (20%)，阐明文章主要假设和理论依据以及你的结果并做出解释 (尤其要说明你做出的结果是否与原文完全相同，不同点在哪，你认为原因是什么)。若能对原文在模型、测量等方面提出异议并加以测验，则为

更佳。另需上交完整数据和可在任何计算机上 R 程序都可直接运行出结果的可重复程序文件 (20%)。报告和可复制文件均需通过邮件提交，邮件名统一为“Method_ 姓名 _ 期末”。提交日期见“课程进度”更新或课上通知。

E-mail 政策

E-mail 为本课师生交流的主要方式。学生来信请以“PoliData_ 姓名:”方式作为邮件标题开头。我会及时回复学生来信。也请诸生秉承尊重彼此时间的精神，避免以下几种情况：

1. 提问教学大纲上已有内容。
2. 提问缺勤课程内容。所缺内容请参见教学大纲和课件。
3. 提问已经缺勤几次，还剩几次免责机会等。
4. 要求延长作业或期末项目时间。学生应按时提交作业和项目。提交时间通常会在大纲中显示或在课上当众公布，如有异议可在期限前提出。但在期限后基于公平原则，不会再作更改或延长。
5. 要求更改分数而没有具体理由。如前所述，本课教学原则之一是保证公平。我希望每位学生都能在同等评价标准下获得其应得的的所有分数。所以欢迎诸位在有异议时申诉分数。基于效率考虑，请参照以下程序进行申诉：
 1. 通过 email 提出申诉的具体内容和具体理由（如，“X 次作业第 X 题正确答案为 XXX，我的答案为 XXX，两者一致，但仍被扣了分数，请问为何？”）。
 2. 我会针对申诉做出回复。如需要当面沟通，我会邀请你来办公室详谈。但如是“[grade grubbing](#)”或要求比别的学生获得更多分数的机会，恕不受理。

课程进度

第 1 周 (2019-09-09): 描述性统计 (Descriptive Statistics)

Gary King. “Replication, Replication”. *PS: Political Science & Politics* 28.03 (1995), pp. 444-452.

Moore and Siegel (2013): Chapter 1;

Wonnacott and Wonnacott (1978): Chapter 2;

陈正昌 and 贾俊平 (2016): 第 1 章.

Bonus: Linear Algebra

第 2 周 (2019-09-16): 概率论 (Probability Theory)

Moore and Siegel (2013): Chapter 9;

Wonnacott and Wonnacott (1978): Chapter 3.

Bonus: Calculus

第 3 周 (2019-09-23): 整体和样本分布 (Population and Sampling Distribution)

Moore and Siegel (2013): Chapter 10-11;

Wonnacott and Wonnacott (1978): Chapter 6.

陈正昌 and 贾俊平 (2016): 第 4 章.

Bonus: Intro to R

[Systematic organizing and naming](#)

第 4 周 (2019-09-30): 假设检验 (Hypothesis Testing)

Wonnacott and Wonnacott (1978): Chapter 8-9.

陈正昌 and 贾俊平 (2016): 第 5-6 章.

Bonus: Rmarkdown

作业 1

第 5 周 (2019-10-07): 国庆假期 (No Class)

(无必读材料)

Bonus: Data Input and Recoding

第 6 周 (2019-10-14): 相关性和独立性 (Correlation and Independence)

Wonnacott and Wonnacott (1978): Chapter 10, 14.

陈正昌 and 贾俊平 (2016): 第 16-18 章.

Bonus: Data Manipulation

提交复刻研究进展报告 1 (备选文章及出处)

提交作业 1

第 7 周 (2019-10-21): 简单回归及推导 (Simple OLS and Estimator Derivation)

Wooldridge (2016): Chapter 2.1-2.2

陈正昌 and 贾俊平 (2016): 第 19 章.

Bonus: Hypothesis Test

第 8 周 (2019-10-28): OLS 估测值性质 (Properties of OLS Estimators)

Wooldridge (2016): Chapter 2.3–2.6

Bonus: Simulation

作业 2

第 9 周 (2019-11-04): 高斯马尔科夫定理 (Gauss-Markov Theorem)

Wooldridge (2016): Chapter 5

Bonus: Binary OLS

第 10 周 (2019-11-11): 多元回归 (Multiple Regression)

Wooldridge (2016): Chapter 3

陈正昌 and 贾俊平 (2016): 第 20 章.

Bonus: Multivariate OLS

提交复刻研究进展报告 2 (备选文章概述：理论、假设、可用数据)

提交作业 2

第 11 周 (2019-11-18): 定理诊断 (Regression Diagnoses)

Wooldridge (2016): Chapter 4, 6.

陈正昌 and 贾俊平 (2016): 第 21–22 章.

Bonus: Diagnosis

第 12 周 (2019-11-25): 共线性与异方差性 (Multicollinearity and Heteroskedasticity)

Wooldridge (2016): Chapter 8, 13, 14.1–14.2.

Bonus: OLS Visualization

第 13 周 (2019-12-02): 自相关性内生性 (Autocorrelation and Endogeneity)

Wooldridge (2016): Chapter 10.1–10.3, 10.5, 12, 15.1–15.3.

Bonus: Fixed Effect and IV Model

作业 3

第 14 周 (2019-12-09): 调节变量与缺失值 (Moderation and Missing)

Thomas Brambor, William Roberts Clark, and Matt Golder. “Understanding Interaction Models: Improving Empirical Analyses”. *Political Analysis* 14.1 (2006), pp. 63-82.

Wooldridge (2016): Chapters 9.5a

Bonus: Interplot and Multiple Imputations

第 15 周 (2019-12-16): 非连续因变量 (General Linear Model)

Wooldridge (2016): Chapter 7.5, 17.1.

Bonus: GLM

提交复刻项目完整报告

提交作业 3

第 16 周 (2019-12-23): 讲座：因果推断简介 (Guest Lecture: Causal Inference)

TBA

Bonus: N/A