**北京大学**

体育经济分析

阅读与写作指导手册

姓 名 ：周正卿

**2022年3月**

**目 录**

[目 录 I](#_Toc1588972451)

[第1章 阅读：引言密码 3](#_Toc1900833512)

[第2章 写作：观察性经验研究的核查清单 7](#_Toc1342103412)

[2.1 流程 7](#_Toc1711301043)

[2.1.1 信度分析 7](#_Toc588145430)

[表2.1 性格色彩内容 12](#_Toc93014869)

[第3章 读写：训练营范文 14](#_Toc2080291914)

[3.1 论文大焖锅 | JEP 卫星遥感影像数据在经济学中应用 14](#_Toc284941791)

[3.2 论文大焖锅 | J\*\* 家庭电脑的使用和人力资本的发展 16](#_Toc128148527)

[3.2.1 研究问题: 16](#_Toc2013678995)

[3.2.2 研究动机（Motivation）: 16](#_Toc1708075892)

[3.2.3 研究设计: 16](#_Toc70123748)

[3.2.4 研究意义（Implication） 16](#_Toc1748794080)

[3.3 论文大焖锅 | AER 比较发展的殖民起源：一项经验调查 18](#_Toc1520909718)

[3.4 论文大焖锅 | AER 比较发展的殖民起源：一项经验调查 20](#_Toc431780185)

[3.2.1 研究问题: 20](#_Toc582326082)

[3.2.2 研究动机（motivation）: 20](#_Toc1071480795)

[3.2.3 研究设计: 20](#_Toc1727341470)

[3.2.4 研究意义（implication） 20](#_Toc1744146144)

[第4章 写作：文献报告内容指南 3](#_Toc712460658)

[1. Introduction/ Motivation 3](#_Toc2104946981)

[2. Background/Literature Review 3](#_Toc198308989)

[3. Proposed Approach 3](#_Toc84557979)

[3.1. Theory 3](#_Toc1679262386)

[3.2. Empirical 3](#_Toc1132832628)

[4. Data Source and Type 4](#_Toc2075448141)

[5. Empirical Results 4](#_Toc480027566)

[6. Conclusion 4](#_Toc1874723630)

[7. Your Remarks 4](#_Toc599980626)

[第5章 写作：研究计划内容指南 5](#_Toc1438658517)

[2.1 Motivation for the Project（研究或项目动机） 5](#_Toc1015313646)

[2.2 Literature Review（文献回顾） 5](#_Toc471389260)

[2.3 Proposed Approach（研究方法和拟解决的关键性问题） 5](#_Toc572119037)

[2.4 Data Source and Type （数据来源和种类） 6](#_Toc1320367240)

[2.5 Data Description（数据的基本描述，前提是已经拿到了所 使用的数据） 6](#_Toc1463678229)

[2.6 Anticipated Results （预期的结果） 7](#_Toc614818418)

[2.7 Potential Topics 8](#_Toc1709325609)

# 阅读：引言密码

(Source:<https://people.ses.wsu.edu/chouinard/wp-content/uploads/sites/285/2015/08/The-Introduction-Formula.pdf>)

When I arrived at UBC, my colleague John Ries, who had been hired the year before, explained to me that Jim Brander had given him a formula for writing introductions. I'm afraid I didn't pay much attention at the time because I thought it would stifle my creative juices (is that a mixed metaphor?). Finally, I think I ended up internalizing the rules and now I thought I should make them explicit because they have served us well and I wish I could referee more papers that follow them.

当我来到UBC时，我的同事John Ries（他在前一年被聘用）向我解释说，Jim Brander给了他一个写引言的公式。恐怕我当时并没有太在意，因为我认为这会扼杀我的创作灵感（这是不是一个混合的比喻？） 最后，我想我最终把这些规则内化了，现在我想我应该把它们明确化，因为它们为我们提供了很好的服务，我希望我可以推荐更多的论文来遵循它们。

看讨论的时候零星看到了两个，问题…还有哪些捏

Hook: Attract the reader's interest by telling them that this paper relates to something interesting. What makes a topic interesting? Some combination of the following attributes makes Y something worth looking at.

Y matters: When Y rises or falls, people are hurt or helped.

Y is puzzling: it defies easy explanation.

Y is controversial: some argue one thing while other say another.

Y is big (like the service sector) or common (like traffic jams).

钩子。吸引读者的兴趣，告诉他们这篇论文与一些有趣的事情有关。是什么让一个话题变得有趣？以下属性的某种组合使Y成为值得关注的东西。

Y天然重要。当Y上升或下降时，人们会受到伤害或帮助。

Y是令人费解的：它不容易解释。

Y是有争议的：有些人争论的是一件事，而其他人说的是另一件事。传统理论不能解释新现象

Y是现实中体量很大（如服务行业）或常见的（如交通堵塞）。

在这部分需要避免的事情包括：

The bait and switch: promising an interesting topic but delivering something else, in particular, something boring.

“诱导转向法”（商业上常见的骗术，商家用低价消费品吸引顾客过来，但转而推销给他们更贵或者更不好的东西）： 向大家许诺一个有趣的话题，但结果讲了一个别的东西，特别是无聊的东西。

"all my friends are doing it" : presenting no other motivation for a topic than that other people have written papers on it.

“我朋友们都在做”：除了说很多人都在做这个话题以外讲不出别的理由。

Question: Tell the reader what this paper actually does. Think of this as the point in a trial where having detailed the crime, you now identify a perpetrator and promise to provide a persuasive case. The reader should have an idea of a clean research question that will have a more or less satisfactory answer by the end of the paper. Examples follow below. The question may take two paragraphs. At the end of the first (2nd paragraph of the paper) or possibly beginning of the second (3rd paragraph overall) you should have the "This paper addresses the question" sentence.

问题。这个部分是关于论文真正要讲的问题。这部分你要做的事打比方来说可能和在一场审判里基于充分的细节证据，对案犯进行指证一样。目的是让读者对论文研究问题有清楚的认识，并且期待在文章结束获得比较满意的答案。这部分可以通常分为两段，而且在两段之间（这里作者的意思应该是第一段末或第二段初的“承上启下句”）你应该说“本文将处理并解决这个问题”。

Antecedents: Identify the prior work that is critical for understanding the contribution this paper will make. The key mistake to avoid here are discussing papers that are not essential parts of the intellectual narrative leading up to your own paper. Give credit where due but establish, in a non-insulting way, that the prior work is incomplete or otherwise deficient in some important way.

“前作回顾”（或者“文献回顾”，英文是Antecedents）：这部分的目的是指明与本文最为相关的关键先前文献，并以此让读者更容易理解本文的贡献。这部分的常见错误是讨论一些仅仅是在同一领域，但和本文主旨并不相关的论文。这部分的注意事项是，你需要给前作足够的地位和贡献认可，但同时，尝试以不失礼貌的方式指出它们在某方面的不完全或者不完善（从而论证本文的改进看起来是有价值的）

Value-Added: Describe approximately 3 contributions this paper will make relative to the antecedents. This paragraph might be the most important one for convincing referees not to reject your paper. A big difference between it and the earlier "question" paragraph is that the contributions should make sense only in light of prior work whereas the basic research question of the paper should be understandable simply in terms of knowing the topic (from the hook paragraph). John suggests that "Antecedents" and "Value-added" may be intertwined. They may also take up to 3 paragraphs.

“增加值”（即“本文贡献”）：列举大约三个本文相比之前文献的贡献。这一段可能对于审稿人决定是否拒掉你的论文至关重要。这一段和“问题”那一段的重要区别是：贡献是基于特定先前文献规定的范畴的，而问题则应该是关于整个话题，让对此有大而化之了解的读者都能理解的。因此从行文关系上，增加值对应前作回顾，而问题则对应引子。作者说James Brander提出文献和贡献有时可以夹杂在一起写，篇幅大概是三段。

Road-map: Outline the organization of the paper. Avoid writing an outline so generic that it could apply to any paper ("the next section is the middle of the paper and then we have the end"). Instead customize the road map to the project and possibly mention pivotal "landmarks" (problems, solutions, results...) that will be seen along the way. But keep this short because many readers will now be eager to get to the heart of the paper.

“路径图”（Road map）：这部分即下文结构概览。这部分应避免写一些看起来对所有文章都适用的没有信息量的话（比如“这部分是文章的中间，接下来我们就到了结尾”）。相反，应尝试这用关键性的“路标”归纳一些量身定制的结构单位，比如“问题、解决方案，结果”(problems, solutions, results…)。这部分总体来说写得越短越好，因为读者接下来通常更有兴趣的是论文主体部分。

Brander suggests that you write the intro first but then read and edit it every time you compose other parts of the paper. Thus by the end, the intro will have received more attention, more times, than any other part of the paper. The introduction is not just important because of the "first impressions" idea that it will tilt the referee for or against you (though it probably will). It is also vital to making sure you know yourself what you are doing in the paper and why. If you can't write a good introduction, then you may be writing the wrong paper.

最后，作者简单说了一下，把引言写好是很重要的。最好的写法是先写引言，然后根据后面内容的增加会随时调整引言。这样论文写完的时候引言就被改了很多次了。写好引言不仅是为了给读者和审稿人更好的第一印象，更重要的是帮自己理解自己的论文到底真正意义在哪里。从某种意义上来说，如果你一直写不出好的引言，那多半是因为你在写一篇错误的论文。

# 写作：观察性经验研究的核查清单

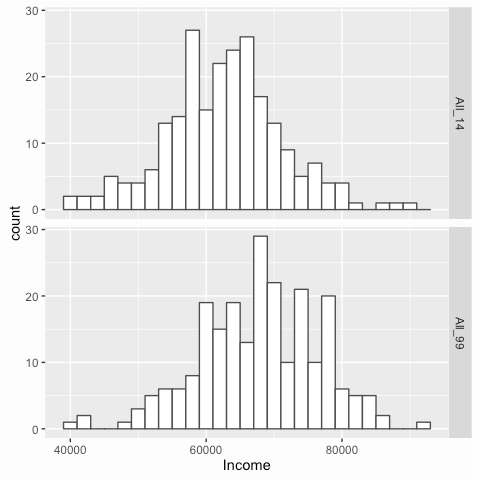
I have been teaching Intro to Statistics and Causal Inference to political science students at the graduate level over the past few years. I find that I was constantly repeating myself on some very basic points. Below is a checklist for analyzing observational data that I hope will be helpful to new practitioners. They are extremely basic and are in no way comprehensive, but in my humble opinion, can remove many potential spurious results in empirical analyses. The bottom line is: please look at your data first!

在过去的几年里，我一直在给研究生阶段的政治学学生讲授《统计学入门》和《因果推理》。我发现我在一些非常基本的观点上不断重复自己。下面是一份分析观察性数据的清单，我希望对新的从业者有所帮助。它们是极其基本的，绝不是全面的，但依我愚见，可以消除经验分析中许多潜在的虚假结果。底线是：请先看一下你的数据!

## 2.1 流程

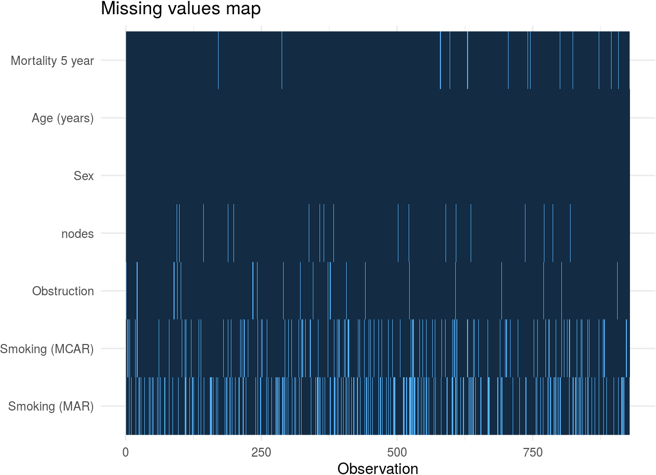
### 2.1.1 信度分析

1. Draw the histograms of the key variables, including the treatment and the outcome. Are these distributions highly skewed or have outliers?画出关键变量的柱状图，包括治疗和结果。这些分布是否高度偏斜或有离群值？

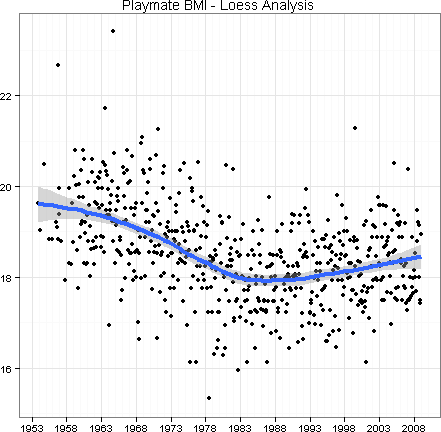


Source: https://afit-r.github.io/histograms

1. Understand the missing data problem in your data by making a plot and think about how to deal with it.通过作图了解你的数据中的缺失数据问题，并思考如何处理。

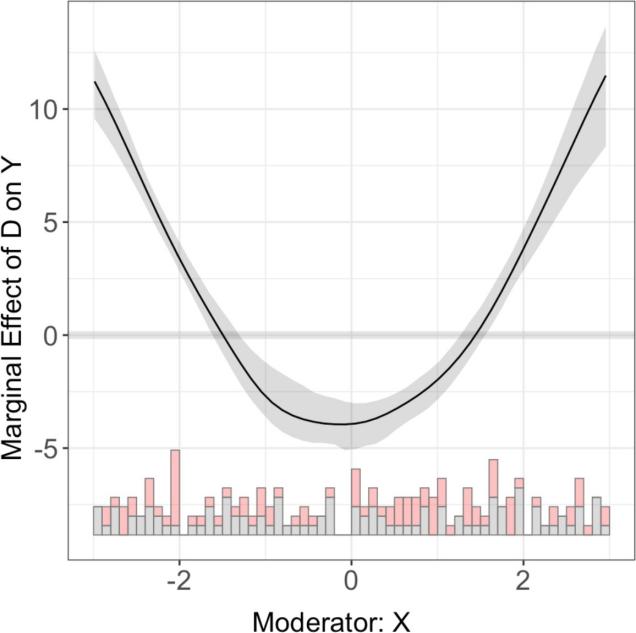


1. Draw a bivariate scatterplot of the treatment and the outcome or a scatterplot between the residualized treatment and residualized outcome. Overlay it with a loess curve. Does your result hold when you “winsor” 5% of the extreme values in your treatment or outcome variables?画一个治疗和结果的双变量散点图，或者画一个剩余治疗和剩余结果的散点图。用黄土曲线覆盖它。当你在治疗或结果变量中 "winsor "5%的极端值时，你的结果是否成立？



Source: https://learnr.wordpress.com/2009/03/10/loess-smoothing/

1. Check whether your “treatment” and “control” groups overlap on key pretreatment covariates.检查你的 "治疗组 "和 "对照组 "是否在关键的治疗前协变量上有重叠。
2. Remember to cluster your standard errors properly—ideally using blocked (wild) bootstrap. One sanity check is to collapse your data to the level where treatment is assigned and run a simple regression using the collapsed data.记住要对你的标准误差进行适当的分组——最好是使用封锁式（野生）把靴法。一个理智的检查是将你的数据折叠到分配治疗的水平，并使用折叠后的数据运行一个简单的回归。
3. If you model includes an interaction term, check whether the linearity assumption looks plausible.如果你的模型包括一个交互项，检查线性假设是否看起来很合理。



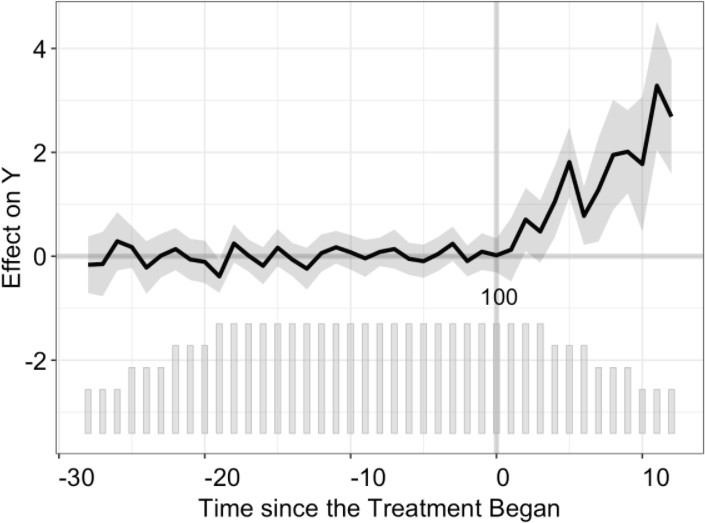
Source: <http://bit.ly/interflex>

1. If you’re analyzing panel data, understanding where your treatment variation comes is crucial. Draw a plot to show how the treatment status changes within a unit over time, for example:如果你在分析面板数据，了解你的治疗变化的来源是至关重要的。例如，画一个图来显示一个单位内的治疗状态如何随时间变化。



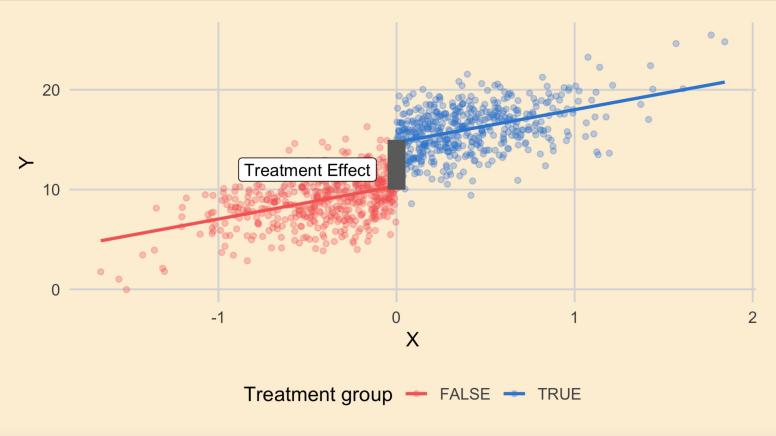
Source: <http://bit.ly/panelView>

1. If you use difference-in-differences design (or use a twoway fixed effects model), draw a dynamic treatment effect plot.如果你使用差分设计（或使用双向固定效应模型），请画出动态治疗效果图。



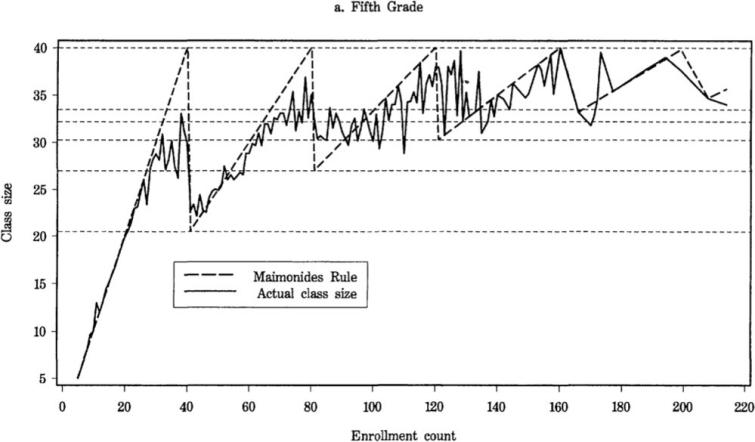
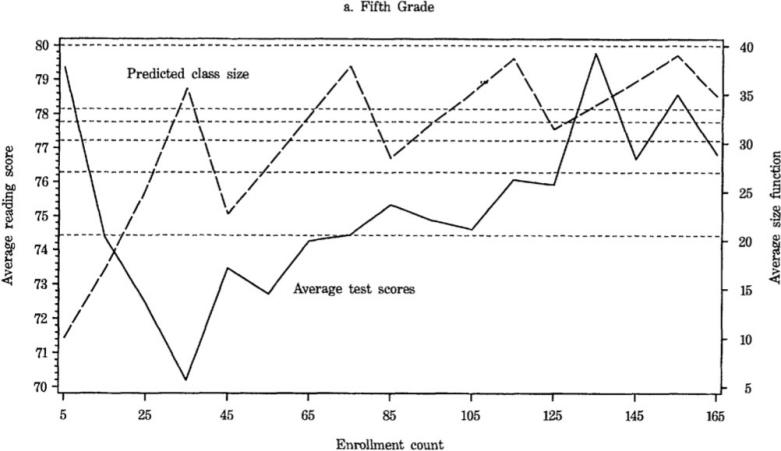
Source: https://bit.ly/2NZZGo7

1. If you use a regression discontiguity (RD) design, draw a RD plot for the reduced form. If it’s a fuzzy RD, draw one for the first stage as well. Same for an interrupted time series design.如果你使用的是断点（RD）设计，请画出还原形式的RD图。如果是模糊的RD，也要为第一阶段画一个。对于中断的时间序列设计也是如此。



Source:https://wytham.rbind.io/html/cbus\_dsmeetup/econ\_toolkit\_2019-06-19.html#54

1. If you use an instrumental variable (IV) design, compare your IV estimates with your OLS estimates. A big discrepancy is suspicious (if your primary concern for the OLS is upward bias) and needs explanation. When your instrument, treatment, and outcome variables are continuous, plotting both the first-stage and the reduced form relationships will be helpful. 如果你使用工具变量（IV）设计，将你的IV估计值与你的OLS估计值进行比较。巨大的差异是可疑的（如果你对OLS的主要担心是向上的偏差），需要解释。当你的工具、治疗和结果变量是连续的，绘制第一阶段和缩小形式的关系将是有帮助的。



Source: https://academic.oup.com/qje/article-abstract/114/2/533/1844228

表2.1 性格色彩内容

|  |  |  |
| --- | --- | --- |
| 色彩 | 优点 | 缺点 |
| 红色 | 追求快乐和自由、开朗热情、善于表达、助人为乐、乐于分享 | 容易激动、缺少分寸、行事随意 |
| 黄色 | 追求成就和目标、坚定自信、目标导向、敢于冒险、求胜心强 | 控制欲强、喜欢批判、缺乏耐心 |

续表 2.1

|  |  |  |
| --- | --- | --- |
| 色彩 | 优点 | 缺点 |
| 蓝色 | 追求严谨和完美、成熟稳重、讲究精确、心思细腻、重视默契 | 敏感多疑、要求苛刻、消极悲观 |
| 绿色 | 追求稳定和平和、与世无争、耐心柔和、稳定低调、先人后己 | 缺乏自信、害怕冲突、缺少主见 |

# 读写：训练营范文

## 3.1 论文大焖锅 | JEP 卫星遥感影像数据在经济学中应用

原文标题：The View from Above Applications of Satellite Data in Economics

推送摘要：Journal of Economic Perspectives最新发表论文简单介绍了卫星遥感影像数据在经济学中应用情况

原文链接：<http://www.ingentaconnect.com/content/aea/jep/2016/00000030/00000004/art00008;jsessionid=gb5q0r9ijli23.x-ic-live-02>

在过去十多年里，伴随计算机技术、工程学以及地理学综合运用，经济学家们对全世界社会和经济状况有了更详细了解。其中，卫星遥感影像数据应用就是上述综合运用重要体现。在学界，研究者利用算法从卫星影像剥离出社会经济特征数据进行相关研究已成为普遍现象。2017年克拉克奖（John Bates Clark）获得者Dave Donaldson和Adam Storeygard教授合作撰写并最近发表在Journal of Economic Perspectives上论文“The View from Above Applications of Satellite Data in Economics”结合数据优势、注意事项、数据源以及存在问题简单介绍了卫星遥感影像数据在经济学中应用情况。

在数据优势方面，作者认为存在以下三点：第一，数据获取较易，且变量较广。这极大降低了研究成本；第二，影像分辨率较高。大部分遥感数据多基于30\*30米栅格单位进行统计，部分甚至为0.5\*0.5米单位统计。这大大丰富经济学者研究，尤其涉及地理空间政策评估方面分析；第三，数据覆盖较广，部分数据收集频率较高，且收集不受当地变故或政策变动影响。尽管卫星遥感数据因上述优势在经济学研究中取得广泛应用，但便捷性更多体现在小范围地理空间研究上。若涉及大范围影像数据应用，研究者则需具备一定计算机处理技术和相关知识。

在研究事项方面，作者强调需注意以下几点：1）卫星轨道。大部分遥感影像均为卫星拍摄。这些卫星主要位于两大轨道：其一是地球同步静止轨道。即卫星始终位于地球表面固定位置，且能够持续对同一地区进行连续观测。但因其距离地球较远，影像分辨率通常较低。其二是太阳同步轨道。即卫星保持固定角度在同一时间经过地球同一位置，多位于南北极上空。因其距离地球高度不一，研究者在使用中需在影像分辨率与观测频率间权衡。2）传感器和光谱。遥感卫星通常安装多个传感器，一个传感器只能基于一种或几种光谱进行数据记录。因不同光谱数据具有不同用途，学者需结合实际情况甄别使用。3）影像处理。遥感影像处理将会影响数据质量，其中主要涉及两种技术。一是正射校正：由于卫星观测地球存在角度偏差，研究者需对原始影像进行经纬度校正；二是多重叠加：云雾等外在因素会对影像造成偏差，学者需利用图像镶嵌技术对影像进行多重叠加处理，从而消除外部随机干扰。4）影像解析和分类。为了便于研究，研究者通常使用两大方法对影像数据进行预先处理。一类是对影像进行解析，并获取研究者想要内容，比如海拔、坡度等数据；另一类是对影像进行分类。比如，将土地按用途进行分类。其中，分类算法主要有两种：一是无监督学习，即研究者对数据不进行预先处理，所有过程均交由计算机进行建模分类；二是监督学习，即研究者对数据先进行主观分类标注，并结合标注对计算机进行训练得到最优模型，然后基于最优模型对原始影像数据进行分类处理。

随后，本文简单介绍了遥感影像相关数据源及经济学应用文献。其中，涉及到的遥感数据有夜间灯光、气候与天气、地形、农业用地与适种作物、城市用地、自然资源、空气污染等。同时，作者强调上述数据综合运用可为研究带来便利。比如，结合夜间灯光数据和日间影像资料，研究者利用迁移学习等机器算法进行解析，最终分类出衡量消费和资产的指标数据。这将极大降低因大范围调查所产生的研究成本。

最后，针对影像数据应用中存在问题，作者分别从数据规模与高维灾难、空间相关、度量偏误以及隐私问题对其展开探讨。具体来说，数据规模和高维灾难主要指影像数据常具有大规模和高维数特征，这将导致传统线性分析方法失效。研究者需结合经济理论选择合适模型进行降维处理。空间相关主要指影像数据常展现较强空间相关性。针对影像数据在研究中被作为自变量或因变量，学者需采取不同措施进行处理。度量偏误主要指在对数据进行预先主观处理时，尽管学界在算法方面取得较大进步，但一二类错误仍会产生。隐私问题主要指高分辨率影像数据将会在一定程度上侵犯个人隐私，研究者需要予以重视。考虑到未来遥感影像数据日益丰富，且容易获得，本文认为这将给经济学多个分支领域带来研究便利。

## 3.2 论文大焖锅 | J\*\* 家庭电脑的使用和人力资本的发展

原文标题：home computer use and the development of human capital

### 3.2.1 研究问题:

Y：弱势家庭儿童与青少年的人力资本发展

（研究了6类被解释变量：1、电脑拥有与使用；2、时间使用；3、学习成绩；4、认知能力与电脑技能；5、非认知型结果（健康状况）；6、父母考虑电脑使用与作业而制定规矩）

X：家用电脑的使用

### 3.2.2 研究动机（Motivation）:

（问题天然重要）

家用电脑使用对儿童人力资本的发展的影响是当前研究的主题，理解家用电脑使用对儿童发展的风险与收益在国内和国际电脑保有量不平等的背景下是非常重要的。大多数研究认为在美国，家用电脑的使用和不同的教育结果之间存在正相关，而近期大部分证据表明家用电脑使用和儿童发展之间取决的混淆因素繁多，因此不可能反映真实的因果效应。

### 3.2.3 研究设计:

断点回归（RD）

### 3.2.4 研究意义（Implication）

该研究**发现（Rfind）**家用电脑的使用对人力资本的发展既有正面影响也有负面影响，通过6类变量分开探究了正面效应和负面效应，对揭示家用电脑的使用对劳动市场变量（如工资）的净效应非常有帮助；阐释了父母在塑造家用电脑的使用对儿童与青少年发展影响中的潜在角色。

1父母被警告严禁作出任何虚报收入的行为，并被要求提供家庭收入的证明文件。2即便如此，仍会可能有家庭试图低报收入，但由于申领项目代金券的收入标准在事前是未知的（取决于申报项目的家庭数目），这样的作弊行为不会影响断点附近研究对象的行为，也就不会影响RDD的估计效果。

## 3.3 论文大焖锅 | AER 比较发展的殖民起源：一项经验调查

原文标题：the colonial origins of comparative development: an empirical investigation

**划线部分的翻译：**frederich a. von hayek（1960）认为，英国的普通法传统比法国的民法要好，后者是在拿破仑时代发展起来的，以限制法官对国家政策的干预（见alseymour m. lipset，1994）。最近，拉斐尔-拉波尔塔等人（1998年，1999年）强调了殖民起源（殖民者的身份）和法律起源对现行制度的重要性，并表明英美法系国家和前英国殖民地拥有更好的财产权和更发达的金融市场。(1998)认为，前英国殖民地比前法国、西班牙和葡萄牙殖民地繁荣，因为他们从英国继承了良好的经济和政治体制及文化。与这种注重殖民者身份的方法相反，我们强调殖民地的条件。特别是，在我们的理论和数据中，重要的不是殖民者的身份或法律渊源，而是欧洲殖民者是否能安全地在某一特定地点定居：在他们不能定居的地方，他们创造了更糟糕的制度。在这方面，我们的论点与斯坦利-恩格曼和肯尼思-索科洛夫（1997）的论点密切相关，他们也强调制度，但将其与要素赋值和不平等挂钩。

在经验上，我们的工作与其他一些试图揭示制度和发展之间的联系的尝试有关，还有tograziella bertocchi和fabio canova（1996）以及robin m. grier（1999），他们研究了作为殖民地对战后增长的影响。而hall和jones（1999）则用离赤道的距离作为社会基础设施的工具，因为他们认为，纬度与 "西方的影响 "相关，而后者会导致良好的制度。另一方面，民族语言的分裂似乎是内生的，特别是这种分裂在欧洲的增长时代几乎完全消失了，因为中央集权的国家和市场出现了（见，例如。eugenj. weber, 1976; benedict anderson, 1983）。从经济上看，这两项研究的问题是，他们的工具可能对绩效产生直接影响。例如，wil-liiam easterly和ross levine(1997)认为，民族语言的分裂可以通过创造政治不稳定来影响绩效，而charles de montesquieu[1748](1989)和最近的david e. bloom和jeffreyd. sachs(1998)以及john gallup等人(1998)则认为气候对绩效有直接影响。如果这些变量确实有直接影响，它们也是无效的工具，不能证明是制度在起作用。我们的方法的优点是，以我们已经控制的变量为条件，100多年前的定居者道德应该对今天的产出没有影响，除了通过它对制度的影响。

3.2.1 研究问题:

y：经济发展

x：制度

3.2.2 研究动机（motivation）:

（问题天然重要）

3.2.3 研究设计:

断点回归（rd）

3.2.4 研究意义（implication）

## 3.4 论文大焖锅 | AER 比较发展的殖民起源：一项经验调查

原文标题：the colonial origins of comparative development: an empirical investigation

### 3.2.1 研究问题:

y：弱势家庭儿童与青少年的人力资本发展

x：家用电脑的使用

### 3.2.2 研究动机（motivation）:

（问题天然重要）

### 3.2.3 研究设计:

断点回归（rd）

### 3.2.4 研究意义（implication）

# 写作：文献报告内容指南

In general, the presentation should follow the structure as follows

## Introduction/ Motivation

* What is the question being addressed?
* Why it is important and interesting for both academics and policy makers?
* 正在解决的问题是什么？
* 为什么它对学术界和政策制定者都很重要和有趣？

## Background/Literature Review

* What does the previous literature say?
* What was missing from the previous literature?
* What is the intended main contribution of the paper?
* 以前的文献说了什么？
* 以前的文献中缺少什么？
* 本文预期的主要贡献是什么？

## Proposed Approach

### Theory

* Is there some underlying economic models?
* How to link this study with these theories?
* 是否有一些基本的经济模型？
* 如何将这项研究与这些理论联系起来？

### Empirical

* What is the identification strategy?
* What is the key assumption for this strategy?
* Who are the control group and treatment group?
* What are the key variables? Dependent Variable, Independent Vari­able and Covariates
* 什么是识别策略？
* 这个策略的关键假设是什么？
* 谁是控制组和治疗组？
* 关键变量是什么？因变量、自变量和协变量

## Data Source and Type

* Where does the data come from?
* What is the unit of observation?
* Is it individual, cross-sectional or panel data, how many observations?
* Why is this good data to use for answering the question?
* Briefly show descriptive statistics tables to provide some useful infor­mation
* 数据从哪里来？
* 观察的单位是什么？
* 是个人、横断面还是面板数据，有多少个观察点？
* 为什么说这是用于回答问题的好数据？
* 简要展示描述性统计表，提供一些有用的信息

## Empirical Results

* Tables or figures of results
* Interpretation of results
* Are there some potential problems with the estimation or interpreta­tion of the results?
* How the authors prove the results are robust?
* 结果的表格或数字
* 结果的解释
* 在估计或解释结果方面是否存在一些潜在的问题？
* 作者如何证明结果是稳健的？

## Conclusion

* brifely summarize the main point
* what could we learn from the paper?
* 归纳出主要观点
* 我们能从该文件中学到什么？

## Your Remarks

* Strongpoint and Weakness?
* What do you learn from the paper?
* 强点和弱点？
* 你从这篇论文中学到了什么？

# 写作：研究计划内容指南

本指南是从分析的角度介绍一篇好的经验研究项目计划书，应该大致 包括哪些内容。而关于研究项目的具体实施，具体内容如下：

## Motivation for the Project（研究或项目动机）

* What is the question being addressed?（拟发现或提出的提出的问题）
* Why is it so important and interesting from a curtain perspective of business?（从研究问题的理论、经验和政策三方面来说为什么有意义? ）

## Literature Review（文献回顾）

* 对之前的文献进行简要的综述，主要结论是什么？
* 这些研究可能存在哪些缺陷？
* 本文的研究以哪些重要文献作为研究基础，并有可能在哪些方面对上 述文献进行扩展或改进？

## Proposed Approach（研究方法和拟解决的关键性问题）

* What are your key variables and specific estimation techniques?你的 项目中最关心的变量是什么？自变量和因变量是什么？在项目要克服 的主要的识别的困难是什么？（即是否样本不随机，存在自选择或者互 为因果等关系？）
* What is your main estimation strategy/method?所以，为了克服上述 困难，你打算所采用的主要经验研究方法是什么？（比如我们上课介绍 的，实验结合多元回归、工具变量和双重差分法等等）为什么这种方 法能够解决你研究中的识别困难？

## Data Source and Type （数据来源和种类）

* Where does the data come from? What is the unit of observation?数 据从何而来，以什么为单位？

1. Is it individual, cross-sectional or panel data, how many observations? 从数据类型上看，属于哪种类型？cross-sectional、panel、experimental
2. Is the data ready to use? If not, what transformation or recoding will you have to do?数据是否可以马上使用？还是说要进行一些处理，比如去掉一些缺失值和异常值。
3. Which items will be used in your data set as the measures of your interested variable? 研究问题中自变量和因变量分别对应数据库中哪个或哪些变量？

* Why is this good data to use? Advantage and disadvantage?为什么你所使用的数据适合做本项目的研究，该数据有哪些优势和劣势？

### Data Description（数据的基本描述，前提是已经拿到了所 使用的数据）

* Descriptive statistics tables （if you already have some descriptive statis­tics, this is great, if not we will wait...）数据信息的基本描述：我们最关心的因变量、自变量和其他控制变量的基本统计性描述信息（均值、方差等等）如果有可能的话可以做一点比较两组的差异分析（类似我们上课做的）,再用散点图大致看一看因变量和自变量之间的关系。

### Anticipated Results （预期的结果）

* What issues will be left unresolved? What is the anticipated contribu­tion of the research?本研究项目期望得到的结果是什么？其研究结果对移民研究和政策有 哪些直接的贡献？研究结果的潜在的局限性有哪些？如何能够克服？

### Potential Topics

•人口与劳动经济学

1）健康和营养方面：比如肥胖、吸烟和喝酒对工资、家庭消费和福利 的影响2）教育和培训：在职培训对城市居民和农民工就业及工资的 影响。3）失业就业研究：“四万亿”投资与就业、“一带一路”策略与 就业、工作流动性研究4）劳动力市场市场化程度研究5）农民工工资 上涨的原因？劳动力短缺还是制度原因（最低工资制度）6）人口问题：计划生育、二胎的影响

•教育经济学

1）城乡中学教育质量的差异？ 2） “补课”是否有用？ 3） “快慢班”对学 生成绩的影响？ 4）男女生成绩的差异

•发展经济学

1. 收入分配：工资不平等和收入不平等相关
2. 城乡移民迁移：

a） 迁移决策的决定：教育、土地、家庭结构及社会网络如何决定移 民的决策？

b） 对迁移地的影响：比如农民工迁移对当地就业和工资的影响

c） 对迁出地的影响：比如农民工迁移对农村收入分配和贫困的影响

d） 其他衍生问题：子女教育、婚姻质量、青年农民工等等

1. 腐败问题：反腐运动对家庭消费、奢侈品消费、餐饮消费的影响？

•增长领域

1）劳动力市场与经济增长：比如劳动力市场发展中国经济增长2）政 治与经济增长：比如官员晋升机制与中国经济增长3）不平等与经济 增长：比如：库兹涅兹曲线在中国4）全要素生产率（TFP）与经济增 长：生产率与中国经济增长

•环境与健康

1）空气、水污染对健康、劳动生产率、消费、幸福感等方面的影响

其实还有很多很多的问题都可以做，上面只是我暂时想到的一些题目，特点是一般都有可能有相应的数据支持，有一定的文献积累。也特别欢迎你们想一些自己感兴趣的题目，我相信：*"*兴趣永远是最好的老师！ *”*。

1.1.1只有模型（实证），没有实证（模型）

1.1.2模型不足

1.1.3实证不足

1.1.4基于全球（单一国家）的，没有单一国家(全球)的

1.1.5基于 case study(系统检验)的，没有系统检验(case study)的

1.1.6基于定性（量）的，没有定量（性）的

1.1.7基于之前的，没有现在

组织方式：将团队成员分成2组，每组下载经济研究最新几期上任一篇论文，最好找参考文献较多且中英文都有引用的论文。参考文献格式按照我发的“参考文献格式”。小组完成后，发给另外一个小组检查反馈并修改，然后由老师提意见并形成最终练习。一般来说，一个学期系统训练1-2次，建议放在学期开始。之后在平时的推送训练中不断提醒学生注意细节。·频度：老师定，每次1个小时。组会作业：可以放在其他组会之后，时间1小时，老师把握时间。组会效果：在任何情况下都注意细节，避免因为细节失分。·

组织方式：可以指定团队中比较 senior的学生作为协调人，现场打开顶刊和领域期刊、NBER工作论文系列，按照兴趣挑选论文。下载后快速阅读提取研究问题以及motivation,老师计时在10分钟内提取出来。之后，老师带领大家讨论 motivation种类。如果不在我列的范围呢，务必告诉我一下。这是一个大事对我来说。·组会作业：把组会最后一个或者半小时空出。由协调人把阅读论文分给学生，学生结合讨论写短推送并讨论后定稿。协调人严格把握时间，在一个小时内定稿。组会效果：通过不断的练习能够掌握不同文献状态下如何 motivate出自己的研究问题，让读者及审稿人觉得你的研究重要。组会效果：通过不断的练习能够掌握不同文献状态下如何 motivate出自己的研究问题，让读者及审稿人觉得你的研究重要。频度：老师定，每次4-5个小时。·

组织方式：可以指定团队中比较 senior的学生作为协调人，现场打开顶刊和领域期刊、NBER工作论文系列，按照兴趣挑选论文。下载后不要阅读摘要，而是快速阅读完前言部分，提取研究问题、motivation、研究设计、发现、稳健性讨论及频度：老师定，每次4-5个小时。·implication部分。每篇论文尽量在20-30分钟内完成，老师计时。之后老师带领大家