

定量研究方法的基础

原理与概念

林景 讲师

南京财经大学法学院

定量研究的初步印象

- 定量研究的概念界定
- 定量研究的资料形态
- 定量研究的哲学基础
- 定量研究的基本结构

定量研究的初步印象

- 定量研究的概念界定
 - 对社会现象的数量特征及其关系进行测算和分析的方法
- 定量研究的资料形态
 - 统计部门汇总的数据、问卷调查采集的数据、实验量表记录的数据、历史文本编码转换的数据
- 定量研究的哲学基础
 - 实证主义：强调以测量和计算的方式获取经验知识的研究取向
- 定量研究的基本结构
 - 确定问题→提出假设→操作化与测量→抽样与采集数据→分析数据→得出结论

定量研究的项目申请书

- <http://www.tiaozhanbei.net> 第十二届国赛入选作品

挑战杯

全国大学生课外学术科技作品竞赛

主办单位：共青团中央 中国科协 教育部 中国社会科学院 全国学联
承办单位：北京航空航天大学

首页 了解挑战杯 挑战杯动态 创业资讯 作品库 2019国赛 省赛/报备 高校专区

人气榜

名次	作品名称	简介
1	面向“十二五”规划的四川省现代服务业现状调查与...	作品主要针对江苏省农村食品消费安全存在的问题，通过定量定性的方法，从农村消费者角度出发，找出农村食品消费安全问题产生的主要原因，从消费者、从业者、食品安全监管部门等多角度、全方位提出政策建议，首次系统构建了江苏省农村食品消费安全防范网络，希望能够提供给卫生监督机构和食品安全监督部门，从理论上指明未来食品安全规制的发展方向，并结合监管部门工作实际，提出可操作性较强的具体措施，解决当前存在的实际问题。
2	公平与效益的博弈——关于中部地区农村中小学撤点并...	食品安全问题是现阶段我国面临的重大公共安全问题之一，农村食品消费安全直接关系到广大农民的生命安全健康权益保障以及社会的稳定和谐，也是从中央到地方各级人民政府重视的现实问题。农村食品消费安全问题是当今社会所关注的现实问题，关注此也是每一个公民社会责任意识的体现，作品极具现实意义和适用指导价值。研究报告采用了整群随机抽样的调研方法，针对江苏农村不同经济社会发展水平，选择了苏南、苏中、...(查看更多)

基本信息

项目名称：农村食品消费安全现状与安全保障研究报告--以江苏省农村食品安全现状为例

来源：第十二届“挑战杯”作品

小类：管理

大类：哲学社会科学类社会调查报告和学术论文

简介：

喜欢 0

详细介绍：

定量研究的项目申请书

对谁进行问卷调查？

科学性、先进性及独
特之处

作品采用定量定性相结合的方法。数据资料主要来源两个方面，一是按照苏南、苏中、苏北的地理位置及相应社会经济发展水平进行食品安全问卷调查；二是对各级卫生监督所、卫生局、食品药品监督管理局等相关部门访谈和资料收集。从消费者、从业者、食品安全监管部门等多角度、全方位提出政策建议，首次尝试系统构建农村食品安全防范网络，提出的观点反映了当前学术界的最新成果，提出的建议有效结合农村食品安全监管部门的工作实际。

收集什么资料？

定量：统计部门汇总的数据

定性：从档案记载中摘录的资料

项目选题从何而来

- 准确了解中国社会的典型事实 (stylized facts)
- 国家级高端智库的调研报告 (25所)
 - 国务院发展研究中心
 - 中国人民大学国家发展与战略研究院
 - 北京大学国家发展研究院
- 具有新闻公信力的媒体调查报道
 - 财新传媒《财新周刊》《中国改革》
 - 广东省出版集团《时代周报》
 - 广州日报报业集团《南方窗》

定量研究可以回答什么问题？

- 是什么 (what) → 描述统计 (descriptive statistics)
- “哥伦布大交换”与中国经济增长
 - 1470-1900年，农业技术进步（南美作物引种）在中国呈现什么面貌？
 - 1470-1900年，中国经济状况是什么样？
- 气候冲击与中国农民起义
 - 25-1911年，中国发生了多少次气候灾害？
 - 25-1911年，中国发生了多少起农民起义？

定量研究可以回答什么问题？

- 是什么 (what) → 描述统计 (descriptive statistics)
- 麦宜生 (Ethan Michelson) : 中国律师业调查
 - 中国执业律师的年龄状况是什么样 (t1 + t7) ?
 - 以中青年为主 (26-50岁) / 以老年为主 (50岁以上)
 - 中国职业律师的性别状况是什么样 (v1 + t7) ?
 - 以男性居多 / 以女性居多
 - 中国什么地区的执业律师最多?
 - 东部发达城市 / 中西部欠发达城市
 - 中国执业律师数量增长最快是哪个时期 (t5 + t7) ?
- 为什么要 “+ t7” ?



定量研究可以回答什么问题？

- 是什么 (what) → 描述统计 (descriptive statistics)
- 卡莱罗纳人口研究中心 (Carolina Population Center)
 - 中国健康与营养调查 (China Health and Nutrition Survey)



定量研究可以回答什么问题？

- 为什么 (why) → 推论统计 (inferential statistics)
- “哥伦布大交换” 与中国经济增长
 - 1470-1900年，农业技术进步（南美作物引种）是否对中国经济有积极影响？
 - 农业技术进步→人口增长→经济繁荣
 - 农业技术进步→技术革命→经济繁荣
 - 农业技术进步→农业商品化→经济繁荣
 - 农业技术进步→贸易扩张→经济繁荣
- 原因 (cause) 与结果 (effect)
 - 请根据中国律师业调查，写下您想研究的问题

定量研究可以回答什么问题？

- 为什么 (why) → 推论统计 (inferential statistics)
- 一因多果、多因一果、多因多果
- 一因一果
 - 1470-1900年，农业技术进步（南美作物引种）在中国带来怎样的影响？
 - 1470-1900年，中国经济受到什么影响？
- “原因的结果 (effects of causes)” 范式
 - 将研究聚焦于两类社会现象，从作为原因的一方出发探讨两者可能存在的因果关联性 (causality)
 - “气候冲击与中国农民起义”
 - “警察与小偷”

什么是好的定量研究问题？

- 遵循苏格拉底式探究方法（Socratic method of inquiry）的传统
 - 究根本问
 - “哥伦布大交换”与中国经济增长→农业技术进步与经济的关系
 - 气候冲击与中国农民起义→国家能力与地方治理的关系
 - 佛罗伦萨美迪奇家族与文艺复兴→社会关系资源在经济与文化发展中的作用
- 可被经验资料检验
 - 口说无凭，拿出证据
- 从小见大，促进知识积累
 - 邓肯的“人口学传统（demographic tradition）”

定量研究的基本概念：总体、样本与抽样方案

- 定量研究的原理：通过样本（sample）反映总体（population）的情况
 - 总体：在理论上明确界定的社会现象的完整集合体
 - 样本：总体中的一部分/子集



定量研究的基本概念：总体、样本与抽样方案

- 社会科学真正重视的研究对象是总体。由于直接研究总体的成本较高和难度较大，研究者通常借由研究总体的一部分即样本，来间接获得关于总体的信息。
- 定量研究的原理即是通过研究随机样本的属性和特征，来间接反映总体的属性和特征。
 - 样本的属性和特征→样本统计量（sample statistic）
 - 总体的属性和特征→总体参数值（population parameter）

定量研究的基本概念：总体、样本与抽样方案

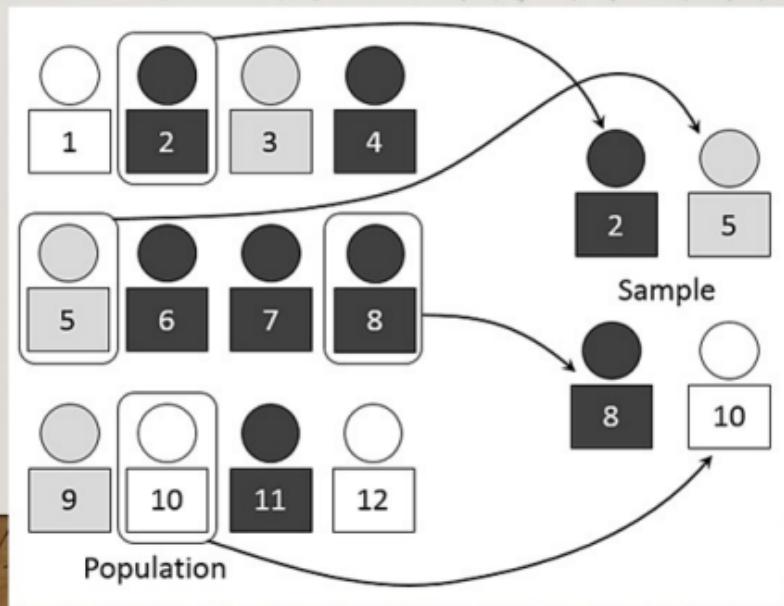
- 样本 (sample) 何以能反映总体 (population) 的情况?
 - 大数定律 (law of large numbers) : 随着样本规模趋于无穷大, 从总体中随机抽取的样本的平均值与总体的平均值之差, 其绝对值将趋近于零。
 - 中央极限定理 (central limit theorem) : 随着样本规模趋于无穷大, 从总体中随机抽取的样本的平均值的分布 (即抽样分布) 将趋近于正态分布。
- 定量研究对样本规模的要求:
 - 小型调查类, 样本规模在100~300之间
 - 中型调查类, 样本规模在300~1000之间
 - 大型调查类, 样本规模在1000~3000之间

定量研究的基本概念：总体、样本与抽样方案

- 抽样：从总体中抽取一定数量的样本作为研究对象。
 - 抽样框：总体中待抽取样本的全部名单。
- 随机抽样：在完整定义的总体中，每一个样本都有一个不为零且相等的中选概率。
 - 简单随机抽样
 - 等距抽样、整群抽样、分层抽样
 - 受访者驱动抽样
- 非随机抽样
 - 偶遇抽样、立意抽样、配额抽样、滚雪球抽样

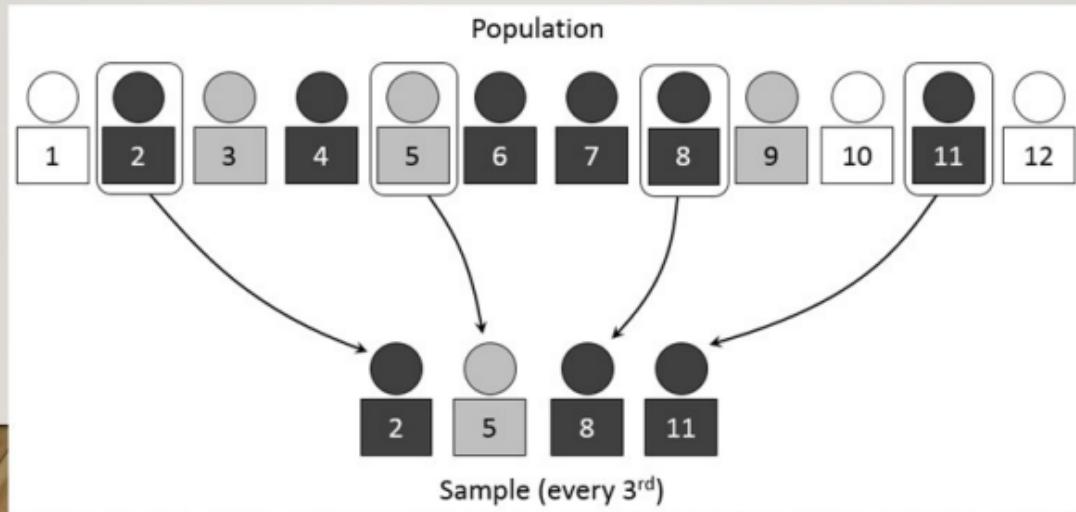
定量研究的基本概念：总体、样本与抽样方案

- 简单随机抽样 (simple random sampling)：
 - 从总体中直接按照一定的比例随机抽取样本，令每一个样本均有同等的中选概率



定量研究的基本概念：总体、样本与抽样方案

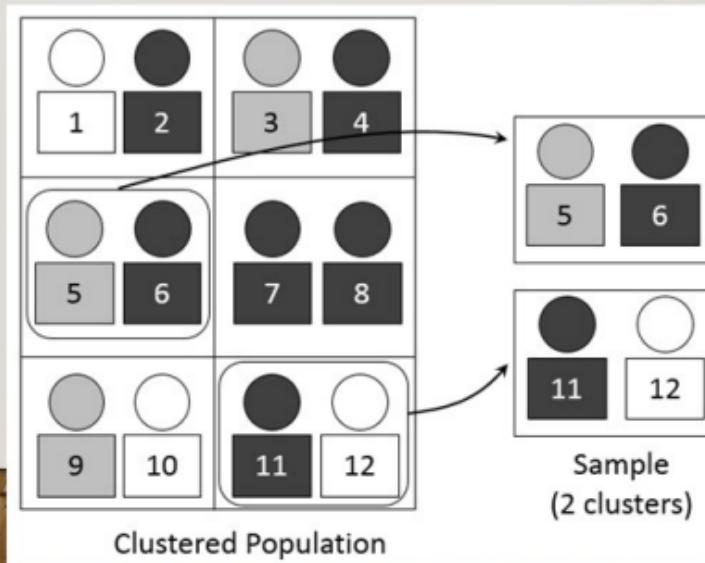
- 等距抽样（equidistant sampling）或系统抽样（systematic sampling）：
 - 将总体中的所有样本按一定的顺序排列，然后根据要抽取的样本量n与总体数量N的比例关系计算抽样距离D ($=N/n$)，接着以简单随机抽样的方式从所有样本中抽出排序为K的一个样本，最后依次按 $K+D$ 、 $K+2D$ 、 $K+3D$ $K+(n-1)D$ 的顺序位置抽取所有样本。



定量研究的基本概念：总体、样本与抽样方案

- 整群抽样（cluster sampling）：

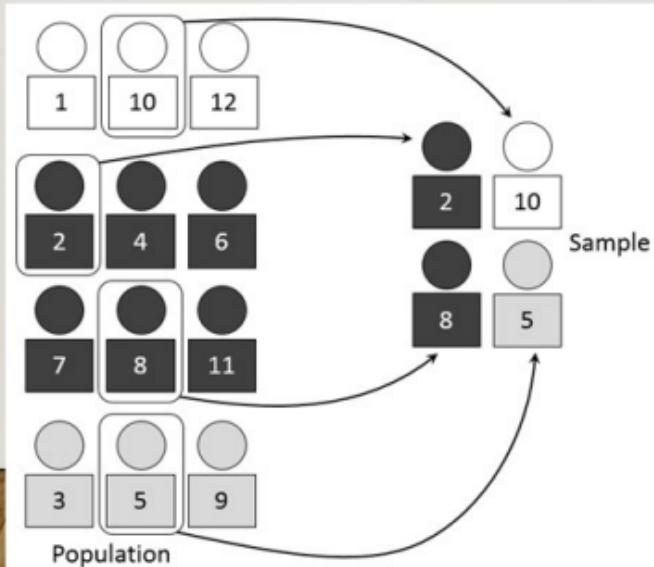
- 先将总体按照一种或几种特征分为几个子群，然后用简单随机抽样的方法从中抽取若干子群，而将这些子群汇总起来就构成总体的样本。



定量研究的基本概念：总体、样本与抽样方案

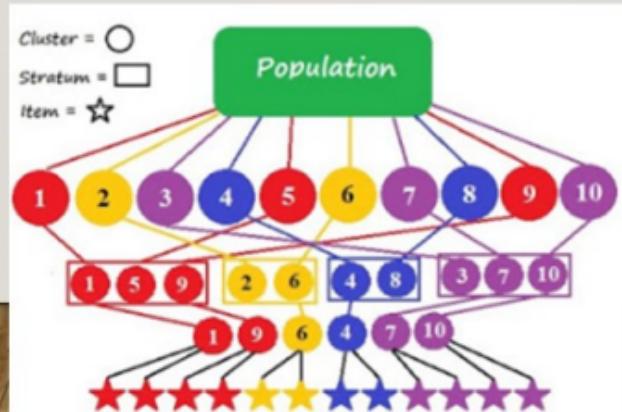
- 分层抽样 (stratified sampling) :

- 先将总体按照一种或几种特征分为几个子总体，每个子总体称为一层，然后从每一层中随机抽取一个子样本，再将其汇总在一起构成样本。



定量研究的基本概念：总体、样本与抽样

- 综合的方法：多阶段抽样（PPS）
 - 先将总体划分成高低不同的层级和将每一层级划分成不同的子群，然后在抽样时按层级从高到低依次在每一层级内随机抽取一定数量的子群，最后将这些不同层级的子群汇总起来即构成总体的一个样本。
 - 大型社会调查项目：全国→省份→地级市→区/县→街道/乡镇→小区/自然村→家户



定量研究的基本概念：总体、样本与抽样方案

- 抽样的一般程序

- 界定总体：在具体抽样前，对从中抽取样本的总体范围和界限做明确的界定；
- 制定抽样框：依据已经明确界定的总体范围，收集总体的全部名单，并对其进行统一编号来建立供抽样使用的抽样框；
- 决定抽样方案：根据不同的研究目的和条件，选择适当的抽样方案；
- 实际抽取样本：严格按照不同抽样方案的执行规则，从抽样框中选取样本；
- 评估样本质量：对样本的质量、代表性和偏差等进行检验和衡量。

定量研究的基本概念：理论假设与操作假设

- 假设
 - 理论假设：从理论直接推演出来，由理论概念构成的待检验命题
 - 操作假设：由可被直接观测的变量构成的、具体的待检验命题
- “哥伦布大交换”与中国经济增长
 - 农业技术进步（南美作物引种）越深入，中国经济增长越快
 - 南美作物引种时间越早的地区，GDP增长越快
 - 有关南美作物引种的技术书籍出版越多的地区，GDP增长越快

定量研究的基本概念：变量、测量及变量类型

- 变量：研究对象的具体特征或属性
 - 操作化：将抽象的概念转化为可观察的具体变量的过程。
 - 测量：按照一定的规则将数值或符号分派于研究对象的特征或属性之上。
- “哥伦布大交换”与中国经济增长
 - 农业技术进步（南美作物的引种）
 - 引种时间先后、范围、规模和产量等
 - 引种技术人员数量、技术著作出版量等
 - 中国经济增长
 - 各行政区的GDP、税役水平、基础建设投入等

定量研究的基本概念：变量、测量及变量类型

- 变量类型：按变量间关系来划分
 - 因变量（dependent variable）：指仅随其他一个或几个变量变动而发生改变的变量。
 - 自变量（independent variable）：指能引起因变量发生改变的变量。



定量研究的基本概念：变量、变量取值及其标签

- 变量类型：按变量的内在性质来划分
 - 定类变量：仅拥有不同类别的属性和特征的变量；
 - 定序变量：属性和特征可以依据一定的逻辑顺序排出等级大小、高低、先后的变量；
 - 定距变量：属性和特征具有相等数量间距的变量；
 - 定比变量：属性和特征可以进行加减乘除运算的变量。

定量研究的基本概念：变量、变量取值及其标签

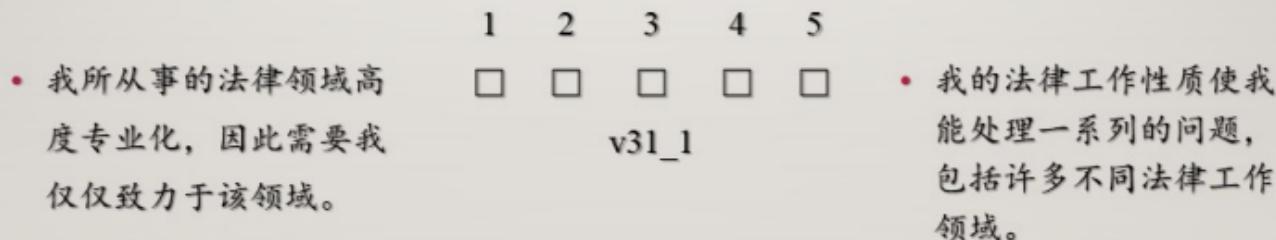
- 定类变量 (nominal variable) 及其取值
- 二分类变量 (binary variable)
 - 虚拟变量 (dummy variable)
 - 性别v1 (男 = 1; 女 = 0) 、是否是仲裁委员会仲裁员v3 (是 = 1; 否 = 0)
 - 其他二分类变量
 - 性别v1 (男 = 1; 女 = 9) 、是否是仲裁委员会仲裁员v3 (是 = 1; 否 = 9)
 - 多分类变量 (categorical variable)
 - 您的政治面貌v5 (群众 = 1; 团员 = 2; 共产党员 = 3; 民主党派 = 4)
 - 您的常住房所有权属于v12 (国家或公有制单位 = 1; 租用 = 2; 私房或商品房 = 3)

定量研究的基本概念：变量、变量取值及其标签

- 定序变量 (ordinal variable) 及其取值
- 您在工作中与其他律师打交道的频率是多少?
 - 您所在事务所的其他律师v22_1:
经常 = 1; 有时 = 2; 很少 = 3; 从不 = 4
 - 在京的非贵事务所的其他律师v22_2:
经常 = 1; 有时 = 2; 很少 = 3; 从不 = 4
- 以下列出几种主要的解决您所受理的“争议解决”类型案件的途径。请问在过去一年中，这些途径各占多大的比例?
 - 通过法庭外的调解或仲裁v30_1:
没有 = 1; 小部分 = 2; 一部分 = 3; 大部分 = 4; 全部 = 5

定量研究的基本概念：变量、变量取值及其标签

- 定距变量 (interval variable) 及其取值
- 不同类型的法律要求不同种类的专业技能.....



定量研究的基本概念：变量、变量取值及其标签

- 定比变量 (ratio variable) 及其取值
 - 您几套住房一共的总价钱 ____ 元 t14;
 - 您几套住房一共的总租金 ____ 元 t15;
 - 您常住房总面积: ____ 使用平方米 (m^2) t16
- 请问邮政编码属于定比变量吗?
 - 您的常住房所在邮政编码: ____ t17



社会学谢宇

2017-9-30 13:28 来自 iPhone 7 Plus

太好了。我终于在已发表的文章中找到定量研究中最愚蠢的错误了—用城市编码算相关系数。而且作者是我们北大的。以后讲定量研究错误有资料，有实货了！

☆ 收藏

107

41

84

定量研究的基本概念：变量、变量取值及其标签

- 不同变量类型的比较

变量类型	变量取值特征	数学性质
定类变量	变量取值为相互排斥、可辨别的类型	=, ≠
定序变量	变量取值之间具有一定的等级序列	>, <
定距变量	变量取值之间具有相等尺度的间距	+,-,
定比变量	变量取值之间可做数量关系的比较，且0点具有意义	+,-, ×, ÷

定量研究的基本概念：变量、变量取值及其标签

- 变量间转换

- 定比变量→定距变量→定序变量→定类变量

- 每月工资收入

- 0元、1050元、2100元.....15000元
 - 0-1000元、1001-2000元.....14001-15000元
 - 低收入、中低收入、中等收入、中高收入、高收入
 - 无收入、有收入

定量研究的基本概念：描述统计与推断统计

- 描述统计
 - 作图：直方图、饼状图、茎叶图、分类图
 - 作表：列联表
- 推断统计
 - 回归分析

案例：船难中的性别、阶级与幸存几率

- “泰坦尼克号”船难事故
 - 20世纪初全世界最大的海上船舶
 - 1912年4月10日首航，从英国南安普顿出发，途径法国瑟堡和爱尔兰皇后镇，驶往美国纽约。14日夜晚，行船在北大西洋与冰山相撞，造成1514名乘客落水遇难，成为近代史上最严重的船难事故之一。



案例：船难中的性别、阶级与幸存几率

- 《泰坦尼克号》电影
- 确定问题
 - 在船难面前，具有什么身份的群体（性别、阶级）更有可能获得救援？



三等舱男性乘客



一等舱男性乘客



三等舱女性乘客



一等舱女性乘客

案例：船难中的性别、阶级与幸存几率

- 《泰坦尼克号》电影
- 提出假设
 - 在船难事故中，女性乘客比男性乘客更有可能获得救援；
 - 在控制其他条件情况下，女性乘客相比男性乘客，幸存几率更高。
 - 在船难事故中，社会经济地位更高的乘客更有可能获得救援；
 - 在控制其他条件情况下，一等舱的乘客比其他舱的乘客，幸存几率更高。
 - 在船难事故中，社会经济地位高的男性乘客比社会经济地位低的女性乘客更有可能获得救援。
 - 在控制其他条件情况下，一等舱的男性乘客相比其他舱的女性乘客，幸存几率更高。

案例：船难中的性别、阶级与幸存几率

- “泰坦尼克号” 乘客数据
 - <https://www.kaggle.com/c/titanic>
- 数据概况
 - 1045名乘客样本
- 变量选择
 - 是否幸存（幸存=1；遇难=0）、性别（男性=1；女性=0）
 - 出发口岸（南安普顿=1；瑟堡=2；皇后镇=3）
 - 船舱等级（一等舱=1；二等舱=2；三等舱=3）
 - 年龄、同行兄弟或配偶数数目、同行父母或子女数目、船票费用



案例：船难中的性别、阶级与幸存几率

- 《泰坦尼克号》电影
 - 不同身份群体的幸存几率比较



案例：船难中的性别、阶级与幸存几率

- 描述统计：不同性别的幸存人数比较、不同船舱等级的幸存人数比较

	遇难	幸存	总数
女性	96 (24.74%)	292 (75.26%)	388
男性	522 (79.45%)	135 (20.55%)	657
总数	618	427	1045

	遇难	幸存	总数
一等舱	103 (36.27%)	181 (63.73%)	284
二等舱	146 (55.94%)	115 (44.06%)	261
三等舱	369 (73.80%)	131 (26.20%)	500
总数	618	427	1045

案例：船难中的性别、阶级与幸存几率

- 推论统计：Logit回归分析

	是否幸存	
	模型1	模型2
性别（参照组：男性）	0.495*** (0.029)	56.933*** (28.989)
船舱等级（参照组：一等舱）		
二等舱	-0.154*** (0.034)	2.051** (0.451)
三等舱	-0.321*** (0.041)	0.654 (0.451)
交乘项		
性别 x 三等舱	- -	0.076*** (0.042)
N	1045	1045
pseudo R-sq	0.323	0.357

注：控制变量包括船票费用、出发地点、同行兄弟姐妹/配偶数目和同行父母/子女数目；模型1系数为偏回归系数，模型2为概率发生比；* p<0.05 ** p<0.01 *** p<0.001