

# 用 R 语言做社会调查抽样设计

蔡苗

2019-04-04



感谢我的家人的支持。

## Acknowledgement

I want to thank my mentor.

# 目录

序	xi
第一章 引言	1
第二章 简单随机抽样	3
第三章 分层抽样	5
第四章 比例数据和回归	7
4.1 比例数据 . . . . .	7
4.2 回归 . . . . .	7
第五章 聚类抽样	9
第六章 非均等概率抽样	11
第七章 大型复杂社会调查	13
第八章 调查对象拒绝回答	15
索引	17



# 表格





# 插图

1.1 这是一个中文标题。 . . . . .	2
-------------------------	---



# 序

在大多数依赖于社会调查进行研究的学科中，最常见遇到的问题就是应该怎么抽样？在那些地方抽样？每个地方抽多少样本？



# 第一章 引言

- **总体 - 参数 (parameter)**: 研究者想要知道但是无法知道的总体特征, 比如全国男性平均身高,
- **样本 - 统计量 (statistic)**: 研究者可以通过抽样进行实际计算的样本特征, 比如全国普查男性平均身高。
- **无限总体**: 总体的量非常大, 我们可以假设我们抽出来的每个观察值都是独立同分布的。
- **有限总体**: 总体的量不大, 我们抽出来的样本并不独立, 但是同分布。

假设我们要研究某大学社会学系总共 30 个班级本科学生的四级考试平均成绩, 此时总体可能无法看做是无限总体。这是因为总共 30 个班级人数较少, 并且不同年级同学四级考试成绩可能会存在系统差异, 因此此总体看做是有限总体比较合理。假设我们要研究某小池塘中的平均汞浓度, 我们可以从池塘的不同边缘抽出 10 小杯样本水, 然后检测每个样本中汞浓度。在此例子中, 我们可以假设总体为无限总体, 因为小池塘中的汞浓度相对均匀, 并且抽取 10 小杯水并不会对总体池塘的水有太大影响。

- 简单随机抽样
- 分层抽样
- 聚类抽样
- 系统抽样

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

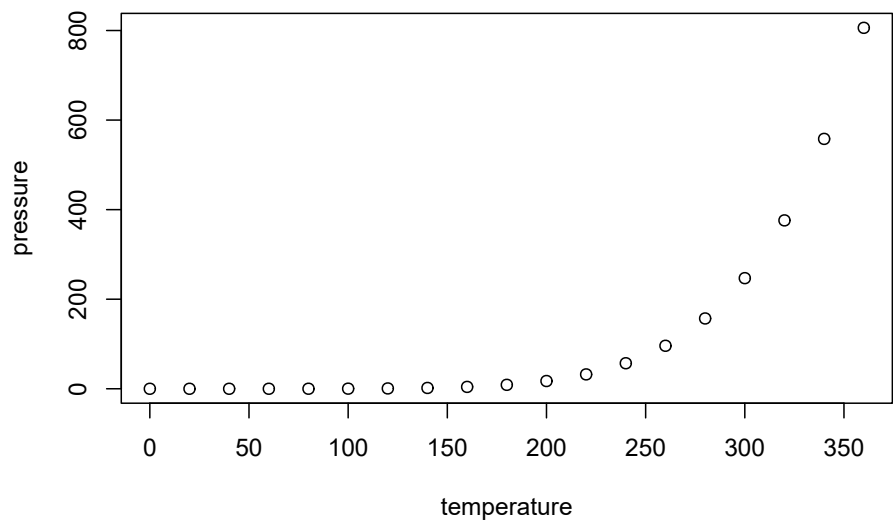


图 1.1: 这是一个中文标题。

## 第二章 简单随机抽样

**简单随机抽样** (Simple Random Sample, SRS): 总体中的不同单位拥有相同的被抽中几率。

假设我们能够从均值为  $\mu$ , 方差为  $\sigma^2$  的总体中不放回抽样得到观测值  $y_1, y_2, \dots, y_n$ , 那么我们可以得到以下关于均数  $\mu$  的点估计量:

$$\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

并且我们可以得到以下性质:

$$E(\hat{\mu}) = \mu V(\hat{\mu}) = \frac{\sigma^2}{n} \frac{N-n}{N-1}$$

第一个性质说明此  $\hat{\mu}$  为无偏无计量。

我们同时还可以得到关于方差  $\sigma^2$  的估计量:

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

并且有以下性质:

$$E(\sigma^2) = \frac{N}{N-1} \sigma^2 \hat{V}(\hat{\mu}) = \frac{s^2}{n} \left(1 - \frac{n}{N}\right)$$





## 第三章 分层抽样



## 第四章 比例数据和回归

### 4.1 比例数据

### 4.2 回归



## 第五章 聚类抽样



## 第六章 非均等概率抽样





## 第七章 大型复杂社会调查



## 第八章 调查对象拒绝回答



# 索引

分层抽样, 1

简单随机抽样, 1

系统抽样, 1

聚类抽样, 1