

基于R与RStudio的心理统计教 学模式探索

南开大学社会心理学系

吕小康

2017/05/21

1 / 35

理念篇

所谓“心理统计”

我以为“心理统计”是个容易产生歧义的用语

所谓“心理统计”

我以为“心理统计”是个容易产生歧义的用语

- 不是一个分支学科，因为不具备方法独特性

所谓“心理统计”

我以为“心理统计”是个容易产生歧义的用语

- 不是一个分支学科，因为不具备方法独特性
- 甚至也很难说具有“领域特殊性”
 - 不学心理统计，只学统计学，也一样能分析心理学数据
 - 只学心理统计，不学其他统计学，很难分析更广阔意义上的数据

所谓“心理统计”

我以为“心理统计”是个容易产生歧义的用语

- 不是一个分支学科，因为不具备方法独特性
- 甚至也很难说具有“领域特殊性”
 - 不学心理统计，只学统计学，也一样能分析心理学数据
 - 只学心理统计，不学其他统计学，很难分析更广阔意义上的数据
- 本质上不过是“统计学在心理学中的应用”

所谓“心理统计”

我以为“心理统计”是个容易产生歧义的用语

- 不是一个分支学科，因为不具备方法独特性
- 甚至也很难说具有“领域特殊性”
 - 不学心理统计，只学统计学，也一样能分析心理学数据
 - 只学心理统计，不学其他统计学，很难分析更广阔意义上的数据
- 本质上不过是“统计学在心理学中的应用”
- 宜理解为一个笼统的“课程名称”，而非精确的“学科名称”

传统心理统计教学的特征之一

传统心理统计教学的特征之一

- 强调纸笔运算
 - 每次看到我的学生还在查阅标准正态分布表、 t 分布表、 F 分布表，我就很心痛！
 - 这完全是在浪费时间！

传统心理统计教学的特征之一

- 强调纸笔运算
 - 每次看到我的学生还在查阅标准正态分布表、 t 分布表、 F 分布表，我就很心痛！
 - 这完全是在浪费时间！
 - 每次看到我的学生在心理统计考试时还在用几十块钱的计算器，我也很心痛！
 - 这完全是在浪费金钱！

传统心理统计教学的特征之二

- “实战性能”有所欠缺
 - 通常要匹配一个所谓的与统计有关的“计算机应用”课程才能完成基本教学任务
 - 而且使用的软件是各种类型的盗版，比如某正则表达式为 $\wedge S.S\{2\}$ 软件

传统心理统计教学的特征之二

- “实战性能”有所欠缺
 - 通常要匹配一个所谓的与统计有关的“计算机应用”课程才能完成基本教学任务
 - 而且使用的软件是各种类型的盗版，比如某正则表达式为 $\wedge S.S\{2\}$ 软件
 - 对于不同类型的数据获得、储存、转换之间欠缺一个统一的处理平台
 - 使得数据自身的兼容性较有局限，对统计软件的路径依赖严重，间接使得人们不愿放弃既有工具而学习新的、更便捷的工具

传统心理统计教学的特征之三

- 可视化呈现较为薄弱
 - 某些研究者似乎认为基于纯数字（如 p 值）的判断比基于图形的判断更“严谨”，更能给出实验效应是否存在、干预结果是否有效等二元化判断的“明确”结论

传统心理统计教学的特征之三

- 可视化呈现较为薄弱
 - 某些研究者似乎认为基于纯数字（如 p 值）的判断比基于图形的判断更“严谨”，更能给出实验效应是否存在、干预结果是否有效等二元化判断的“明确”结论
 - 传统写作时代的绘图工具较为欠缺，开发新的图形、进行数据展示缺少现实工具，且费时费力，故教材编辑者也不愿意从事相关工作

传统心理统计教学的特征之四

- 处理对象多为非文本的量化数据
 - 对文本数据的统计处理较为薄弱，这自然是时代的局限所致
 - 文本抓取、匹配、切割及相关统计描述与模型建立的工作还未进入基础统计学的课程视野

传统心理统计教学的特征之四

- 处理对象多为非文本的量化数据
 - 对文本数据的统计处理较为薄弱，这自然是时代的局限所致
 - 文本抓取、匹配、切割及相关统计描述与模型建立的工作还未进入基础统计学的课程视野

也许还可以有之五、之六、之七.....

改变，从教学/教师/教材开始

一、建立“大统计学”的视野

- 需要推动心理学教学培养单位重新审视所谓“心理统计学”的教学要求
 - 不能完全局限于“传统心理”的范畴来讨论心理学中的统计教学问题
 - 不能完全依赖于传统的“概率论与数理统计”来完成基础统计教学的工作
 - 摒弃“心理统计”小视野，拥抱真正的“统计学视野”！
- 所以
 - 为避免打击面过大，请听现场发言，谢绝秋后算账。

改变，从教学/教师/教材开始

二、引入新一代的统计分析工具

- 比如R与RStudio

改变，从教学/教师/教材开始

二、引入新一代的统计分析工具

- 比如R与RStudio
- 比如Python

改变，从教学/教师/教材开始

二、引入新一代的统计分析工具

- 比如R与RStudio
- 比如Python
- 比如其他软件

三、变革心理学专业统计教育模式

传统模式

- 《概率论与数理统计》（公共课，有些学校还可能没有，比如我们.....）
- 《心理统计》（专业课）
- 《计算机操作》（专业课）
- 研究实践（看导师要求.....）

三、变革心理学专业统计教育模式

可试验的替代性模式

以四大主干课程，贯穿心理学专业的本科统计教育过程，形成体系化的“大心理统计”课程。

- 《数据分析基础》（2学分）
- 《描述性统计》（2学分）
- 《推论性统计》（2学分）
- 《统计应用与结果报告》（2学分）

三、变革心理学专业统计教育模式

可试验的替代性模式：具体说明

- 《数据分析基础》（公共课或专业必修课，2学分，大一上或下）
 - 内容：介绍数据处理基础知识，如数据获取、清洗、制表、制图等，并简单介绍对文本数据的简介处理
 - 此课程中就引入计算化工具，如R
 - 其实在其他类似课程中也可强行植入这些工具！

三、变革心理学专业统计教育模式

可试验的替代性模式：具体说明

- 《数据分析基础》（公共课或专业必修课，2学分，大一上或下）
 - 内容：介绍数据处理基础知识，如数据获取、清洗、制表、制图等，并简单介绍对文本数据的简介处理
 - 此课程中就引入计算化工具，如R
 - 其实在其他类似课程中也可强行植入这些工具！
- 《描述性统计》（公共课或专业必修课，2学分，大一下或大二上）
 - 介绍传统描述性统计的内容，并介绍初步的数据可视化知识
 - 从一开始就介绍使用包括R在内的开源软件进行数据分析、报告撰写、资料共享和团队协作的开放式、团队式教学科研理念

三、变革心理学专业统计教育模式

可试验的替代性模式：具体说明

- 《推论性统计》（专业必修课，2学分，大二上或下）
 - 介绍抽样分布、置信区间、假设检验、线性模型等内容。
 - 研究型院校可要求在此之前完成概率论内容修习，或融合概率论内容与此课程中。
 - 注重利用实际科研和生活化案例与数据进行展示。

三、变革心理学专业统计教育模式

可试验的替代性模式：具体说明

- 《推论性统计》（专业必修课，2学分，大二上或下）
 - 介绍抽样分布、置信区间、假设检验、线性模型等内容。
 - 研究型院校可要求在此之前完成概率论内容修习，或融合概率论内容与此课程中。
 - 注重利用实际科研和生活化案例与数据进行展示。
- 《统计应用与结果报告》（专业必修课，2学分，大三）
 - 介绍综合性、实践性的统计操作与数据分析专题
 - 结合说明APA的写作格式与报告规范，并介绍的文字写作+数据分析+可视化呈现的综合写作平台（如RStudio）。
- 还需要与所在院校的公共教学部门商议综合化的培养方案
- 硕博期间再开设真正偏向心理学专业领域的中高级统计应用专题必修或选修课程

实例篇

以下我将用R与RStudio来展示如何进行课堂教学。这些内容已在南开大学周恩来管理学院各类型、各专业的统计课程得到实践。

时间有限，这里只展示两个教学案例。

示例一：关于数据处理

示例二：关于抽样分布

我的实践，从个案角度证明，大一、大二左右的学生（不仅仅局限于心理学学生，还包括社会学、政治学、行政管理、社会工作等一系列非经济、金融、商科类的社会科学学生），通过开源软件来了解、认识和掌握统计学的基本概念，并不困难，至少是局部可行的。

案例一：基于tidyverse系列包的数据处理流程

这里展示一个刚刚在南开大学社会心理学系大一学生的《社会调查方法》上讲述的例子。此例同时作为研一学生的课程作业。

本例内容来自使用南开大学社会心理学系汪新建教授教育部重大攻关项目《医患信任关系建设的社会心理研究》课题组自编的《中国医患社会心态问卷：患方卷》的试调查工作。该项目的进展请见我们项目组的[GitHub](#)页面。

2017年上半年，我代授大一的《社会调查方法》课程，故以课程作业的形式，五一期间要求选课学员以每人至少10份的调查量，对周围的“患者”进行方便取样和滚雪球取样，利用问卷星系统进行手机在线调查。

- 此处“患者”的操作化定义为30周岁以上、或曾生育过孩子的中国公民。
- 问卷中对“患者”有更详细的操作化界定。

下载Excel格式后我们将部分基础信息提取保存，并将变量名改为英文。

示例一：基于tidyverse系列包的数据处理流程

编制者在试调查之前已找少部分被试进行了预调查，估计较为认真作答的被试，一般应至少有10分钟左右方可填完问卷。被试的回答时间集中于20~30分钟以内，除了极少数老年被试，一般回答时长不会超过60分钟。

鉴于上述情况，研究者认为，此次网络试调查问卷的有效问卷标准如下：

1. 问卷回答时长在10~60分钟之间（包含端点值）；
2. 排除医务人员被试；
3. 排除学生被试。

示例一：基于tidyverse系列包的数据处理流程

编制者在试调查之前已找少部分被试进行了预调查，估计较为认真作答的被试，一般应至少有10分钟左右方可填完问卷。被试的回答时间集中于20~30分钟以内，除了极少数老年被试，一般回答时长不会超过60分钟。

鉴于上述情况，研究者认为，此次网络试调查问卷的有效问卷标准如下：

1. 问卷回答时长在10~60分钟之间（包含端点值）；
2. 排除医务人员被试；
3. 排除学生被试。

根据上述要求，要求学生完成以下问题：

1. 根据上述三个指标，计算有效问卷率；
2. 统计被试的地理来源分布信息。

示例一：基于tidyverse系列包的数据处理流程

问题（1）：有效问卷率

基本思路：

- 剔除医务人员被试和学生被试比较简单，实际上在问卷星平台中就已经加以剔除，这里只是将这一部分被试简单排除即可。
- 重点在于如何提取时间信息
 - 原因在于网站提供的数据中，时间（time2）包含了中文单位（秒）
 - 这正好提供了如何展示进行文本数据相关操作的机会

示例一：基于tidyverse系列包的数据处理流程

问题（1）：有效问卷率

第一步：提取time2一列中的数字信息，即删掉“秒”字，以便后续转换计算。

输入命令：

```
library(readxl) # 用于读取Excel文件
library(tidyverse) # 用于变量操纵
library(stringr) # 用于文本数据操纵
PD <- read_excel("PDSurveyBasic.xlsx") # 读入原始数据
count(PD) # 计算样本量
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    697
```

示例一：基于tidyverse系列包的数据处理流程

问题（1）：有效问卷率的计算

```
time <- str_extract(PD$time2, "[:digit:]+") %>%  
  as.numeric  
time[1:10]
```

```
## [1] 42 3130 1559 1540 926 30 65 965 2955 8
```

- 这里就已经用到了正则表达式（regular expression）。如果没有《数据分析基础》这种课程打底，理解起来较为困难，但是结果仍是直接干脆、一目了然的。
 - 软件操作的妙处之一，就在于可以“不知其所以然而知其然”，这其实是激发进一步探索心理的一种有效方式！
- 同时这也使用了管道函数%>%。

示例一：基于tidyverse系列包的数据处理流程

问题（1）：有效问卷率的计算

由于原数据计时为秒，不太直观，下面转换为分，同时按前述3个条件筛选有效个案。

```
PDclean <- mutate(PD, mins = round(time / 60)) %>%  
  filter(mins <= 60,  
         mins >= 10,  
         relationship > 1,  
         student != 1)  
count(PDclean)
```

```
## # A tibble: 1 × 1  
##       n  
##   <int>  
## 1   450
```

```
round(count(PDclean)/count(PD), digits = 2)
```

```
##       n  
## 1 0.65
```

示例一：基于tidyverse系列包的数据处理流程

问题（1）：有效问卷率的计算

由于原数据计时为秒，不太直观，下面转换为分，同时按前述3个条件筛选有效个案。

```
PDclean <- mutate(PD, mins = round(time / 60)) %>%  
  filter(mins <= 60,  
         mins >= 10,  
         relationship > 1,  
         student != 1)  
count(PDclean)
```

```
## # A tibble: 1 × 1  
##       n  
##   <int>  
## 1   450
```

```
round(count(PDclean)/count(PD), digits = 2)
```

```
##       n  
## 1 0.65
```

如此就得到了有效问卷率为65%。在此基础上，可进一步统计有效问卷中的被试回答时长及分布，为后续大规模调查中筛选被试提供更具参考的依据。

21 / 35

示例一：基于tidyverse系列包的数据处理流程

问题（2）：被试的地理来源分布（基于有效问卷）

思路分析

- 原ip变量中，包含IP数字和真实地理分布信息。
- 可分为以下步骤进行数据处理
 - 提取()中的内容，剔除其余内容；
 - 新生成两个变量，一为province（省份），一为city（省会城市或地级市），这两个信息均在ip一列中可以找到，且用“-”号连接。
- 这些都涉及文本数据处理，并同样涉及正则表达式的运用。

示例一：基于tidyverse系列包的数据处理流程

问题（2）：被试的地理来源分布（基于有效问卷）

```
ip.location <- str_extract(PDclean$ip, "(?<=\\(\\.*(?=\\))") %>%
  str_split("-", n = 2, simplify = TRUE) %>%
  as_tibble %>%
  transmute(province = .[[1]], city = .[[2]])
head(ip.location, 4)
```

```
## # A tibble: 4 × 2
##   province      city
##   <chr>      <chr>
## 1  内蒙古  呼和浩特
## 2   西藏    拉萨
## 3   西藏    拉萨
## 4   云南    丽江
```

学生若能了解"(?<=\\(\\.*(?=\\))"的含义，整串命令的功能就不难掌握了。

- (?<=\\(\\.*(?=\\))表示匹配第一个左括号(之后的内容
- (=\\))表示匹配第一个右括号)之前的内容</li- .*表示上述两个位置之间的任意字符
- str_split("-", n = 2, simplify = TRUE)表示将得到的列以-为分隔符进行切割，分存为两列
- 后续命令表示对新生成的列进行命名。

23 / 35

示例一：基于tidyverse系列包的数据处理流程

问题（2）：被试的地理来源分布（基于有效问卷）

有了`ip.location`这个数据框，后续的统计分析就简单了。

```
ip.location %>% group_by(province) %>%  
  count %>%  
  arrange(desc(n))
```

```
## # A tibble: 26 × 2  
##   province      n  
##   <chr> <int>  
## 1    天津     70  
## 2    辽宁     43  
## 3    山东     33  
## 4    新疆     32  
## 5    云南     29  
## 6    北京     27  
## 7    四川     26  
## 8    重庆     26  
## 9    河南     24  
## 10   黑龙江    21  
## # ... with 16 more rows
```

如此就得到有效样本的省域分布信息（按人数降序排列）。为节省时间，这里只展示省域分布，未对地级市分布做出统计。

24 / 35

示例二：关于抽样分布

- 抽样分布（sampling distribution）的基础知识：
 - 标准定义
 - 抽样分布是样本统计量的分布。这显然精确而“无用”，即不能帮助人真正理解什么是抽样分布。
 - 实质是重复抽样的假想前提下形成的一个统计推论框架，它在现实中是不一定存在的。

示例二：关于抽样分布

- 抽样分布（sampling distribution）的基础知识：
 - 标准定义
 - 抽样分布是样本统计量的分布。这显然精确而“无用”，即不能帮助人真正理解什么是抽样分布。
 - 实质是重复抽样的假想前提下形成的一个统计推论框架，它在现实中是不一定存在的。
 - 准确意思
 - 抽样分布是对同一总体，做出相同样本容量的、重复(无限)多次的简单随机抽样取遍样本统计量的所有可能值后所体现出来的取值规律性。
 - 对这一规律性，采用概率论的基本知识加以描述，即可概括为某一分布（distribution），也即

$$F(x) = P(X \leq x)$$

示例二：关于抽样分布

- 抽样分布（sampling distribution）的基础知识：
 - 标准定义
 - 抽样分布是样本统计量的分布。这显然精确而“无用”，即不能帮助人真正理解什么是抽样分布。
 - 实质是重复抽样的假想前提下形成的一个统计推论框架，它在现实中是不一定存在的。
 - 准确意思
 - 抽样分布是对同一总体，做出相同样本容量的、重复(无限)多次的简单随机抽样取遍样本统计量的所有可能值后所体现出来的取值规律性。
 - 对这一规律性，采用概率论的基本知识加以描述，即可概括为某一分布（distribution），也即
$$F(x) = P(X \leq x)$$
 - 如果能够找到 $F(x)$ 的精确数学表达形式，后续的统计推论即可基于这一概念框架而得到概率意义上的精准推导。

25 / 35

抽样分布的教学难点

1. 理论框架的“非现实性”

- 现实中的研究通常只能有一次抽样，不可能对同一总体进行反复抽样从而得到关于样本统计量的精确分布的直观感知

抽样分布的教学难点

1. 理论框架的“非现实性”

- 现实中的研究通常只能有一次抽样，不可能对同一总体进行反复抽样从而得到关于样本统计量的精确分布的直观感知

2. 现实研究并非总以随机抽样的形式获得样本

- 心理学中更常见的方式反而是随机化实验，即对被试进行随机化分组。
- 但是对这种实验得到的结果，传统教材仍然采用随机抽样的理论框架介绍抽样分布，使得学生有时会产生困惑：明明不是抽样，遑论重复抽样，那此时的抽样分布从何而来？
- 有时我们还常常对观测数据进行统计推论，既无严格意义上的抽样，也无随机化的分配。

抽样分布的教学难点

1. 理论框架的“非现实性”

- 现实中的研究通常只能有一次抽样，不可能对同一总体进行反复抽样从而得到关于样本统计量的精确分布的直观感知

2. 现实研究并非总以随机抽样的形式获得样本

- 心理学中更常见的方式反而是随机化实验，即对被试进行随机化分组。
- 但是对这种实验得到的结果，传统教材仍然采用随机抽样的理论框架介绍抽样分布，使得学生有时会产生困惑：明明不是抽样，遑论重复抽样，那此时的抽样分布从何而来？
- 有时我们还常常对观测数据进行统计推论，既无严格意义上的抽样，也无随机化的分配。

3. 某些统计量的抽样分布数学推导较为困难

- 毕竟，我们不都是数学家，甚至完全没有这种天赋
- 况且，能够找到精确数学形式的分布总是少见的。很多统计量本身就是很难找到精确分布，然而推论总是要做下去.....

建立经验感知的方式：模拟

抽样分布的建立需要一对互相联系的概念：总体（population）与样本（sample）。

不妨以这样的思路进行教学：

建立经验感知的方式：模拟

抽样分布的建立需要一对互相联系的概念：总体（population）与样本（sample）。

不妨以这样的思路进行教学：

- 先从假想的理论分布（如正态分布、二项分布、指数分布等）总体进行重复抽样，模拟某一简单样本统计量的分布，并与数学推导的结果进行对比解释；

建立经验感知的方式：模拟

抽样分布的建立需要一对互相联系的概念：总体（population）与样本（sample）。

不妨以这样的思路进行教学：

- 先从假想的理论分布（如正态分布、二项分布、指数分布等）总体进行重复抽样，模拟某一简单样本统计量的分布，并与数学推导的结果进行对比解释；
- 到假想的理论分布总体进行重复抽样，模拟某一很难或无法从数学推导获得精确分布的样本统计量的抽样分布；

建立经验感知的方式：模拟

抽样分布的建立需要一对互相联系的概念：总体（population）与样本（sample）。

不妨以这样的思路进行教学：

- 先从假想的理论分布（如正态分布、二项分布、指数分布等）总体进行重复抽样，模拟某一简单样本统计量的分布，并与数学推导的结果进行对比解释；
- 到假想的理论分布总体进行重复抽样，模拟某一很难或无法从数学推导获得精确分布的样本统计量的抽样分布；
- 再到从实际的、不满足特定分布的总体进行重复抽样，模拟对应的样本统计量的分布，验证数学推导的结果是否能够应用于现实，并理解数学推导的局限性与模拟的自由性

举例：样本均值的分布

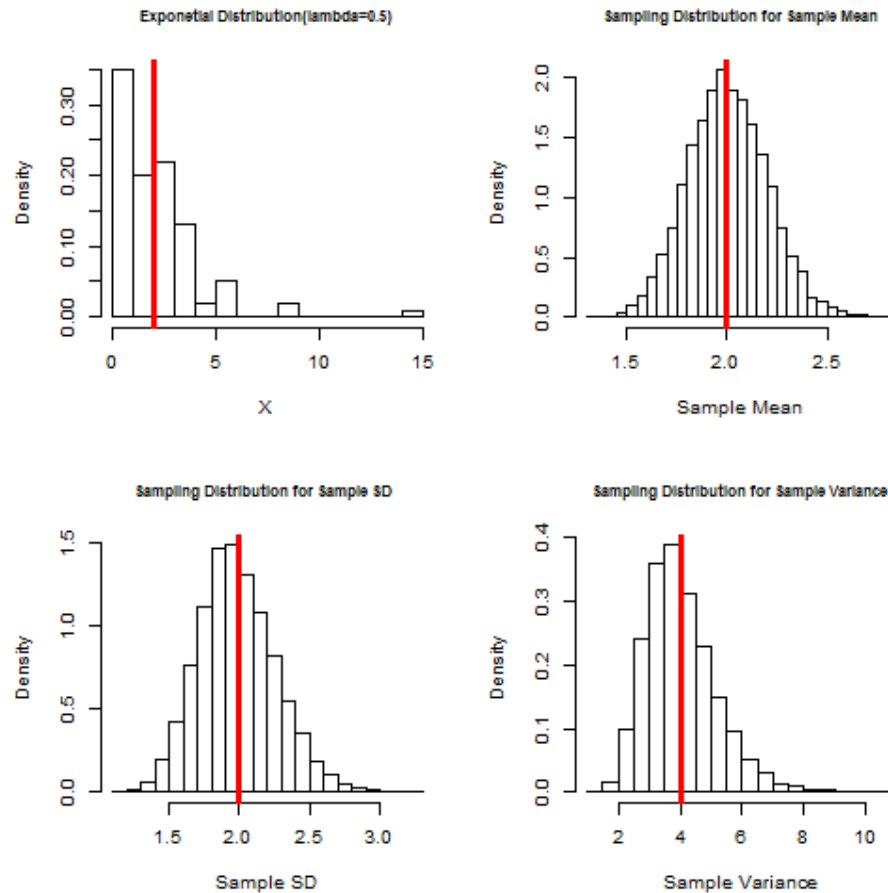
已知总体

$$X \sim \text{Exp}(0.5)$$

，即服从某指数分布，注意该分布本身是右偏的，且是一无限总体。现从中抽取样本量为100的样本，重复10000次，每次计算如下样本统计量，再绘制这些样本统计量的直方图，即可在一定程度上展示该些样本统计量的抽样分布形状，以便形成直观的感知。

- 样本均值（sample mean）
- 样本标准差（sample standard deviation）
- 样本方差（sample variance）

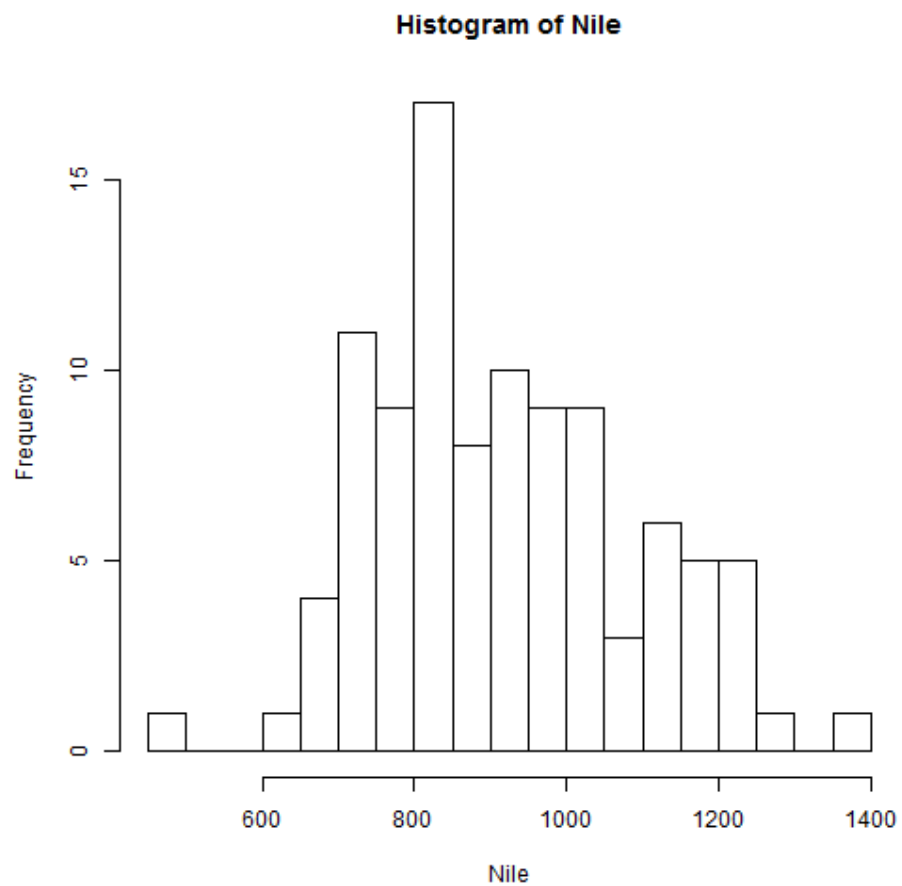
下面该R出场了！



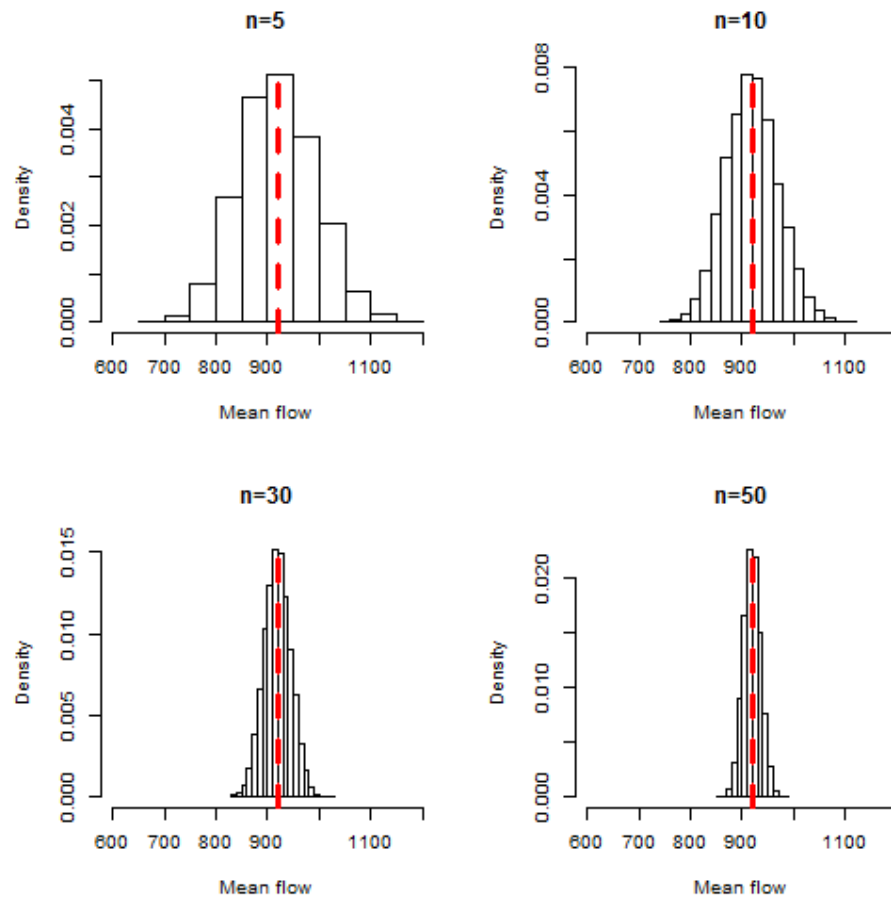
左上图为给定总体的概率密度图，右上图为样本均值的抽样分布示意图；左下图为样本标准差的抽样分布示意图，右下图为样本方差的抽样分布示意图。红色虚线为虚线 29 / 35

Nile数据是R自带的数据库，记录了尼罗河在埃及阿斯量(Ashwan)地区1871-1970年这100年间的年流量值。这一数据服从什么特定的精确分布吗？--不清楚。

```
hist(Nile, breaks = 30)
```



30 / 35



红色虚线表示“总体”均值所在的位置。这是不是就是中心极限定理（Central Limit Theorem）告诉我们的道理呢？

Nile数据，样本量为30、其他样本统计量的抽样分布

样本最小值、样本最大值、样本中位数、样本第一四分位数的抽样分布示意图。红色虚线分别表示“总体”最小值、最大值、中位数和第一四分位数所在的位置。

教师只需在课堂讲解说明并提供一个图形的原始代码，即可要求学生仿照此代码，自行绘制其他图形，并作为作业进行布置。如此可加深学生对抽样分布形成过程的直观认识。

教师只需在课堂讲解说明并提供一个图形的原始代码，即可要求学生仿照此代码，自行绘制其他图形，并作为作业进行布置。如此可加深学生对抽样分布形成过程的直观认识。

实际上关于抽样分布还有一些基于RStudio的shiny平台搭建的动态呈现模式，例如 [Nicole Radziwill](#) 就制作了相关的简单结果展示。这可以作为课堂教学的参考。

如果教师本人精力允许，可以带领学生自行制作相关网页。这样收获更大。

教师只需在课堂讲解说明并提供一个图形的原始代码，即可要求学生仿照此代码，自行绘制其他图形，并作为作业进行布置。如此可加深学生对抽样分布形成过程的直观认识。

实际上关于抽样分布还有一些基于RStudio的shiny平台搭建的动态呈现模式，例如 [Nicole Radziwill](#) 就制作了相关的简单结果展示。这可以作为课堂教学的参考。

如果教师本人精力允许，可以带领学生自行制作相关网页。这样收获更大。

实际上，有了重复抽样情形下的抽样分布及模拟，教师还可利用R进行自助分布（Bootstrap Distribution）、随机化分布（Randomization Distribution）等理论分布的模拟，如此可将推论的背景框架推广至其他观测数据或实验数据的情形。

其他一些观点

- 尽早让学生接触数据分析的基本流程
 - 在没有概率论支撑的情况下就要引入数据分析的基本理念，培养面向数据的分析思维
- 尽早、全面地引入计算工具，深度参与统计教学
 - 我认为无法证明先纸笔计算、再计算机操作的教学模式能够让学生的学习变得更加“扎实”（要不然你给我找出个随机化的现场实验来？）
- 强调可视化在数据分析中的作用
- 对于科研型院校，可适当强调程序思维和编程操作的基础性地位
- 强调统计结果表达的规范性，与国际规范接轨
- 建立统一的、开放的、可编辑的数据展示与教学安全平台，对于节约教师的精力有很大的作用(GitHub可以作为一个有效的平台)。

致谢

- 本幻灯片由谢益辉的 R 包 **xaringan** 生成。
- 部分代码源自与我的研究生及夏骁凯等同学的探讨。
- 感谢 R 语言年会的组织者与志愿者的辛勤服务。
- 其他未尽事宜，相关建议批评，请发信至xkdog@126.com交流探讨。
- 本文档（RAC2017Kahn.Rmd）的相关资料，存放于本人的**GitHub**目录中，可自由分享。