

1.1 试验的意义

王正明 易泰河

系统工程学院 军事建模与仿真系

2019 年 11 月 8 日

引言

“试验” vs “实验”

- 语义区别：“科学实验”与“技术试验”
- “作战实验”与“作战试验”
- Montgomery: Each **experimental run** is a **test**.
An **experiment** is a **test** or series of **runs**.
- 统计领域一般用“试验设计”
- 除课程名字与本次课程外，统一用“试验”一词

教学目的

- ① 掌握因果关系的概念
- ② 掌握试验设计的基本术语
- ③ 了解试验设计发展的大致脉络
- ④ 认同学习实验设计与分析课程的意义

1.1 试验的意义

1.1.1 试验与因果关系

1.1.2 科学研究的范式

1.1.3 试验设计的发展历史

1.1.4 实验设计与分析软件

(1) 什么是统计学？

- 研究如何采集数据、如何从数据中获取知识.



- 实验设计与分析是统计学的重要分支.
- 数据类型:
 - 定量数据: 计量数据、计数数据;
 - 定性数据: 属性数据、有序数据.

(1) 什么是统计学？

- 陈希孺在《数理统计学简史》中指出
 - “正概率”是由原因推结果，是概率论；
 - “逆概率”是由结果推原因，是数理统计。
- 什么是因果关系？
- 直观理解：某种条件下某件事情（原因）引发了令一件事情（结果）。

(2) 什么是因果关系?

关于因果关系的共识

确定**解释变量** A 引起**响应变量** B 的因果关系, 须:

- (1) 相关条件, 即 A 和 B 必须是相关的;
- (2) 时序条件, 即 A 的变化发生在 B 的变化之前;
- (3) 不可替代性, 即 A 与 B 之间的关系不存在替代性解释.

变量之间的因果关系是不能通过演绎推理得到的, 只能通过数据归纳得到.

(2) 什么是因果关系?

- 相关不一定有因果, 有因果一定相关;
 - 啤酒与尿布
- 忽略时序性可能造成因果倒置;
 - 警察数量越多的地区, 犯罪事件数量也越多.
- 注意潜在变量的影响.
 - 健康工人效应: 放射性物质对人体的健康有很大的伤害, 但是铀矿工人的平均寿命却不比常人短.

(2) 什么是因果关系?

- 相关不一定有因果, 有因果一定相关;
 - 啤酒与尿布
- 忽略时序性可能造成因果倒置;
 - 警察数量越多的地区, 犯罪事件数量也越多.
- 注意潜在变量的影响.
 - 健康工人效应: 放射性物质对人体的健康有很大的伤害, 但是铀矿工人的平均寿命却不比常人短.

(2) 什么是因果关系?

- 相关不一定有因果, 有因果一定相关;
 - 啤酒与尿布
- 忽略时序性可能造成因果倒置;
 - 警察数量越多的地区, 犯罪事件数量也越多.
- 注意潜在变量的影响.
 - 健康工人效应: 放射性物质对人体的健康有很大的伤害, 但是铀矿工人的平均寿命却不比常人短.

(3) 研究方法分类

- **观察研究**: 被动的数据搜集方式, 研究人员只观察和记录. 产生的数据称为**观测数据**或**观察数据**.
- **实验研究**: 主动产生数据的方式, 研究者会改变某些变量并观察结果的变化. 产生的数据称为**实验数据**.
- Fisher: 只要遵守良好的试验设计原则, 就可以根据特定的研究得出一般性的结论.

(3) 研究方法分类

- **观察研究**: 被动的数据搜集方式, 研究人员只观察和记录. 产生的数据称为**观测数据**或**观察数据**.
- **实验研究**: 主动产生数据的方式, 研究者会改变某些变量并观察结果的变化. 产生的数据称为**实验数据**.
- **Fisher**: 只要遵守良好的试验设计原则, 就可以根据特定的研究得出一般性的结论.

(4) 什么样的试验可以归纳因果关系?

Example 1.1 (猪鞭虫疗法实验)

克罗恩病是一种慢性肠炎, 含有猪鞭虫卵的饮料可以有效缓解克罗恩病引发的腹部疼痛、出血和腹泻等症状. 开展一项临床实验: 令克罗病患者服用含有猪鞭虫卵的饮料并测量反应. 结果表明症状减轻了, 能否明确猪鞭虫疗法有效?

- 安慰剂效应: 源于患者对医生的信任和对治疗的期望.

双盲实验:

- 将患者随机分为两组, 一组接受猪鞭虫饮料治疗, 另一组则接受安慰剂治疗.
- 患者和负责记录的医生都不知道患者接受哪种治疗.

(4) 什么样的试验可以归纳因果关系?

Example 1.1 (猪鞭虫疗法实验)

克罗恩病是一种慢性肠炎, 含有猪鞭虫卵的饮料可以有效缓解克罗恩病引发的腹部疼痛、出血和腹泻等症状. 开展一项临床实验: 令克罗病患者服用含有猪鞭虫卵的饮料并测量反应. 结果表明症状减轻了, 能否明确猪鞭虫疗法有效?

- **安慰剂效应:** 源于患者对医生的信任和对治疗的期望.

双盲实验:

- 将患者**随机**分为两组, 一组接受猪鞭虫饮料治疗, 另一组则接受安慰剂治疗.
- 患者和负责记录的医生都不知道患者接受哪种治疗.

(4) 什么样的试验可以归纳因果关系?

Example 1.1 (猪鞭虫疗法实验)

克罗恩病是一种慢性肠炎, 含有猪鞭虫卵的饮料可以有效缓解克罗恩病引发的腹部疼痛、出血和腹泻等症状. 开展一项临床实验: 令克罗病患者服用含有猪鞭虫卵的饮料并测量反应. 结果表明症状减轻了, 能否明确猪鞭虫疗法有效?

- **安慰剂效应**: 源于患者对医生的信任和对治疗的期望.

双盲实验:

- 将患者**随机**分为两组, 一组接受猪鞭虫饮料治疗, 另一组则接受安慰剂治疗.
- 患者和负责记录的医生都不知道患者接受哪种治疗.

(5) 随机比较实验

- 双盲实验是随机比较实验的一种：
 - 对试验对象随机分组, 各组在各方面都相似;
 - 确保除了处理方式不同外, 不存在其它潜在变量;
 - 响应变量的差异是处理方式的效应所致.
- 讨论: 随机比较实验能够保证因果关系成立的三个条件吗?
- 良好试验设计原则: 重复、区组和随机化
- 随机抽样服从“机会法则”, 当试验对象较多, 组间的机会变异性(chance variation) 就比较小.

(5) 随机比较实验

- 双盲实验是随机比较实验的一种：
 - 对试验对象随机分组, 各组在各方面都相似;
 - 确保除了处理方式不同外, 不存在其它潜在变量;
 - 响应变量的差异是处理方式的效应所致.
- 讨论: 随机比较实验能够保证因果关系成立的三个条件吗?
- 良好试验设计原则: 重复、区组和随机化
- 随机抽样服从“机会法则”, 当试验对象较多, 组间的机会变异性(chance variation) 就比较小.

(5) 随机比较实验

- 双盲实验是随机比较实验的一种：
 - 对试验对象随机分组, 各组在各方面都相似;
 - 确保除了处理方式不同外, 不存在其它潜在变量;
 - 响应变量的差异是处理方式的效应所致.
- 讨论: 随机比较实验能够保证因果关系成立的三个条件吗?
- 良好试验设计原则: 重复、区组和随机化
- 随机抽样服从“机会法则”, 当试验对象较多, 组间的机会变异性(chance variation) 就比较小.

Example 1.2 (Yule-Simpson's Paradox)

研究某一疾病的某种治疗方式是否有效, 数据如下

		存活人数	死亡人数	存活率
总体	处理组	200	200	50%
	对照组	160	240	40%
男性	处理组	180	120	60%
	对照组	70	30	70%
女性	处理组	20	80	20%
	对照组	90	210	30%

该实验能否归纳出因果关系? 它的设计有什么不合理的地方?

(6) 试验设计的基本概念

- 称衡量试验效果的特征量为**响应**, 也称**指标**, 用字母 y 来表示.
 - 定性响应和定量响应.
 - 单响应试验和多响应试验.
 - 多响应试验通常也转化为单响应试验来处理.
- 称响应的测量值与真值之间的偏差为**误差**
 - 误差是一个随机变量, 一般假定 $\varepsilon \sim N(0, \sigma^2)$.

(6) 试验设计的基本概念

- 称衡量试验效果的特征量为**响应**, 也称**指标**, 用字母 y 来表示.
 - 定性响应和定量响应.
 - 单响应试验和多响应试验.
 - 多响应试验通常也转化为单响应试验来处理.
- 称响应的测量值与真值之间的偏差为**误差**
 - 误差是一个随机变量, 一般假定 $\varepsilon \sim N(0, \sigma^2)$.

(6) 试验设计的基本概念

- 凡对响应可能产生影响的原因都称为因子, 也称因素.
- 试验因子: 根据试验目的选中的可控的因子
 - 定性因子、定量因子、动态因子
- 噪声因子: 试验因子以外的所有因子
 - 随机噪声因子、非随机噪声因子、研究者未意识到的噪声因子
- 称试验因子在试验中所处的状态 (或所取的值) 为水平, 试验因子的水平组合构成处理.

(6) 试验设计的基本概念

- **试验单元**: 一次试验中试验对象、实验员、工具以及材料等组成的基本单元.
- 试验单元之间的差异可能是造成结果波动的原因之一.
- 试验单元、因子和响应共同构成试验的三要素:



Example 1.2: Yule-Simpson's Paradox

研究某一疾病的某种处理方式是否有效, 数据如下

		存活人数	死亡人数	存活率
总体	处理组	200	200	50%
	对照组	160	240	40%
男性	处理组	180	120	60%
	对照组	70	30	70%
女性	处理组	20	80	20%
	对照组	90	210	30%

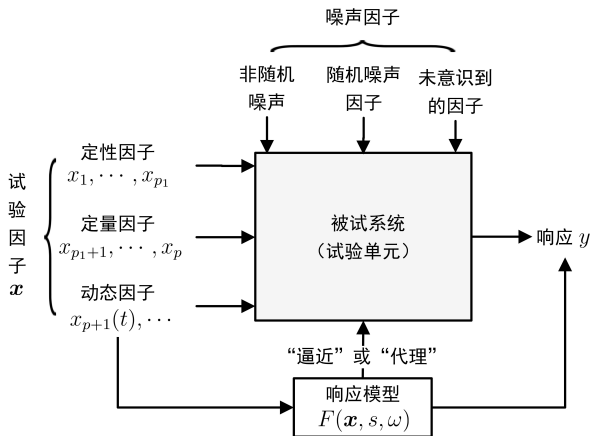
- 响应、因子、水平、处理、试验单元分别是什么?

(6) 试验设计的基本概念

- 称因子与响应之间的数量关系为**响应函数**;
- 响应函数就是因果关系的定量描述;
- 称响应函数的数学模型为**响应模型**;
- 假定存在一个真实的响应函数, 试验的最终目标是为了探究响应函数.

小结

- 遵守良好的试验设计原则可以归纳因果关系.



1.1 试验的意义

1.1.1 试验与因果关系

1.1.2 科学研究的范式

1.1.3 试验设计的发展历史

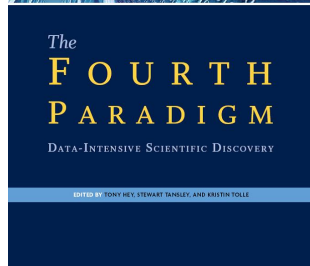
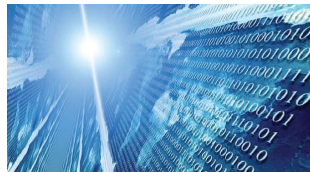
1.1.4 实验设计与分析软件

科学研究四类范式: 经验归纳, 理论推演, 仿真模拟, 数据密集型科学发现.

Science Paradigms

- Thousand years ago:
science was **empirical**
describing natural phenomena
- Last few hundred years:
theoretical branch
using models, generalizations
- Last few decades:
a **computational** branch
simulating complex phenomena
- Today: **data exploration** (eScience)
unify theory, experiment, and simulation
 - Data captured by instruments or generated by simulator
 - Processed by software
 - Information/knowledge stored in computer
 - Scientist analyzes database/files using data management and statistics

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



- 经验归纳中的“经验”来自于观察和试验;
- 理论推演需要试验验证;
- 仿真模拟属于一种新的试验方式.

大数据时代最大的转变, 就是放弃对因果关系的渴求, 取而代之关注相关关系. 也就是说, 只要知道“是什么”, 而不需要知道“为什么”.

——维克托·迈尔, 舍恩伯格: 《大数据时代》

- (1) 大数据工程需要学哪些知识?
- (2) 不需要知道为什么, 实验还有意义吗?

- 经验归纳中的“经验”来自于观察和试验;
- 理论推演需要试验验证;
- 仿真模拟属于一种新的试验方式.

大数据时代最大的转变, 就是放弃对因果关系的渴求, 取而代之关注相关关系. 也就是说, 只要知道“是什么”, 而不需要知道“为什么”.

——维克托·迈尔, 舍恩伯格: 《大数据时代》

- (1) 大数据工程需要学哪些知识?
- (2) 不需要知道为什么, 实验还有意义吗?

Example 1.3 (童便疗法)

中医中有服用“童便”治疗顽固性腹泻的方法, 临床观察很有效. 这究竟是什么原因, 让人很难理解. 研究表明, 顽固性腹泻破坏了肠道的正常菌群, 服用“童便”可补充“双歧杆菌”这一类肠道益菌群, 从而治愈顽固性腹泻. 制成补充双歧杆菌的制剂, 就不会有人去服用“童便”了.

1.1 试验的意义

1.1.1 试验与因果关系

1.1.2 科学研究的范式

1.1.3 试验设计的发展历史

1.1.4 实验设计与分析软件

起源阶段: 1918 – 1940s 的农业实验

- 起源于上世纪 20-30 年代 Fisher 在 Rothamsted 农业实验站的工作.
- Fisher 在上世纪上半叶建立了统计学和试验设计的理论基础, 第二章因子设计的内容主要源自于 Fisher 的工作.
- 第二次世界大战也对统计学的发展产生了重要影响, 序贯分析法就起源于统计学 Wald 分析短程火炮射击精度的工作.

第一个工业时期: 1951 – 1970s 的后期

- Box 与 Wilson(1951) 提出响应曲面法, 回归设计 (第三章) 的理论基本成熟
- 试验设计在化工与流程工业 (process industries) 中的应用, 标志着试验设计应用浪潮的兴起.

第二个工业时期: 1970s 的后期 – 1990

- 由一些企业的质量改进计划推进.
- 美国统计学家 W Edward Deming 将统计质量控制引入日本, 开启了日本统计质量革命时期.
- 由统计质量控制与试验设计产生了管理技术持续质量改进 (CQI) 和全面质量管理 (TQM) 技术, 日本工业产品质量开始得到全面提升.
- 田口玄一利用正交阵列独立提出了一系列部分因子设计的技术, 还提出了鲁棒参数设计和过程鲁棒性的概念.

现代时期: 全球化带来的经济竞争加剧驱动

- 1990 前后, 摩托罗拉工程师 Bill Smith 提出的一种新的持续质量改进方法六西格玛法 (Six Sigma) 逐渐开始流行.
- 随着计算机仿真技术的不断发展, 计算机试验设计逐渐成为问题, (Santer, 1985) 首次提出计算机试验设计的概念.
- *Technometric* 上刊登的论文几乎半数以上与计算机试验设计有关.

1.1 试验的意义

1.1.1 试验与因果关系

1.1.2 科学研究的范式

1.1.3 试验设计的发展历程

1.1.4 实验设计与分析的相关软件

- 商业软件

- SAS, JMP, SPAA, MATLAB, S-PLUS, Minitab, Design-Expert, ...;

- 推荐使用开源的、免费的 R 语言及其编译器 RStudio.

- 简单易学, 添加包 “swirl” 一步步地指导初学者学习;
- 统计学和机器学习领域的研究者在发表论文时, 经常会同时发布一个相应的 R 添加包;
- 能够绘制很漂亮的统计图形, 对学术研究非常有益;
- 本课程中的几乎所有方法都能找到相应的添加包.

总结

- 1.1.1 试验可以归纳因果关系; 试验三要素: 因子、响应、试验单元; 响应函数定量刻画因果关系, 试验探究响应函数.
- 1.1.2 经验归纳中的“经验”来自于观察和试验, 理论推演需要试验验证, 仿真模拟属于一种新的试验方式, 大数据时代因果关系依然重要.
- 1.1.3 试验设计在科技、工业、经济的发展历史和现在都发挥这重要作用.

试验将决定美军通过优化联合部队实现
未来目标的程度.

——《联合作战构想 2020》

试验战略是美国军事转型战略的基础.

——《2001 年四年防务评估报告》

假如有 10% 的工程师使用试验设计方法,
产品的质量和数量都会得到很大提高.

——G. E. P. Box

不懂试验设计的工程师只能算半个工程
师.

——Taguchi

思考

- 利用一架精度为 σ 的天平称 4 个不同的物体, 能否给出一种精度高于 σ 且只需称 4 次的称重方案?