



北京大学  
PEKING UNIVERSITY

---

# 实验物理中的统计方法

第零章：课程介绍

杨振伟

[yangzw@pku.edu.cn](mailto:yangzw@pku.edu.cn)

北京大学物理学院

---

# 课程信息

---

## ➤ 上课时间地点

周四下午7-9节：  
理教302

## ➤ 课程考核

期中考试（闭卷）： 约30%  
期末考试（闭卷）： 约40%  
平时作业与课堂表现： 约30%

## ➤ 先修要求

微积分、线性代数

# 参考书目

---

	作者	书名	出版社	出版年
参考书	Glen Cowan	Statistical Data Analysis	Clarendon Press, Oxford	1998
	茆诗松	概率论与数理统计教程 (第三版)	高等教育出版社	2019
	G. Casella, R. Berger	Statistical Inference	Duxbury	2002
	Frederick James	Statistical Methods in Experimental Physics	World Scientific	2006
	E. T. Jaynes	概率论沉思录	人民邮电出版社	2024
	E. T. Jaynes	Probability Theory — The Logic of Science	Cambridge	2003

# 几个问题

---

- 什么是概率？
- 什么是统计学？什么是统计方法？
- 如果实验得到一系列观测值，如何最有效地从中提取出我们感兴趣的信息？
  - 在实验观测之前如何设计实验？
- 如何对同一个物理量有两个或多个测量结果，如何将这些结果“平均”起来？
  - 假设测量某个长度，两个测量结果分别为  $x_1 \pm \Delta x_1$  和  $x_2 \pm \Delta x_2$ （假设  $x_1 < x_2$ ），平均结果  $\bar{x}$  是多少？可能小于  $x_1$  或大于  $x_2$  吗？
  - 假设实验结果是“显著性”，如何合并？

# 什么是统计学

---

- **Statistics** is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of **data**. (*Wikipedia*)
- **Statistics**, as a subject, provides a body of principles and methodology for designing the process of **data** collection, summarizing and interpreting the **data**, and drawing conclusions or generalities. (R. Johnson et al., *Statistics: Principles and Methods*)

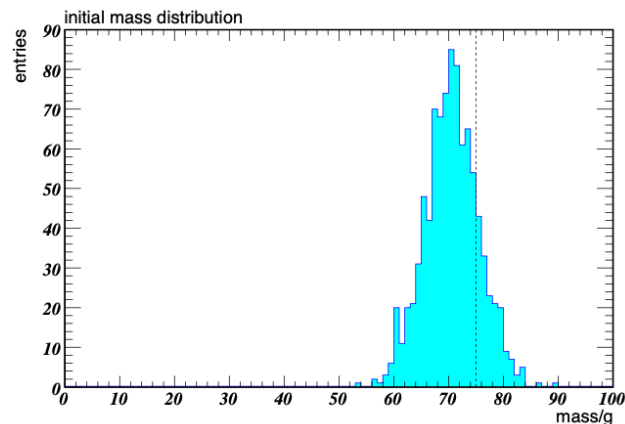
统计学是采集数据、分析数据、展示数据和诠释数据的一门学科。

# 统计描述与统计推断

## ➤ 描述性统计学

- 内容：收集、整理、展示数据，描述性分析
- 目的：描述数据特征，找出数据的规律

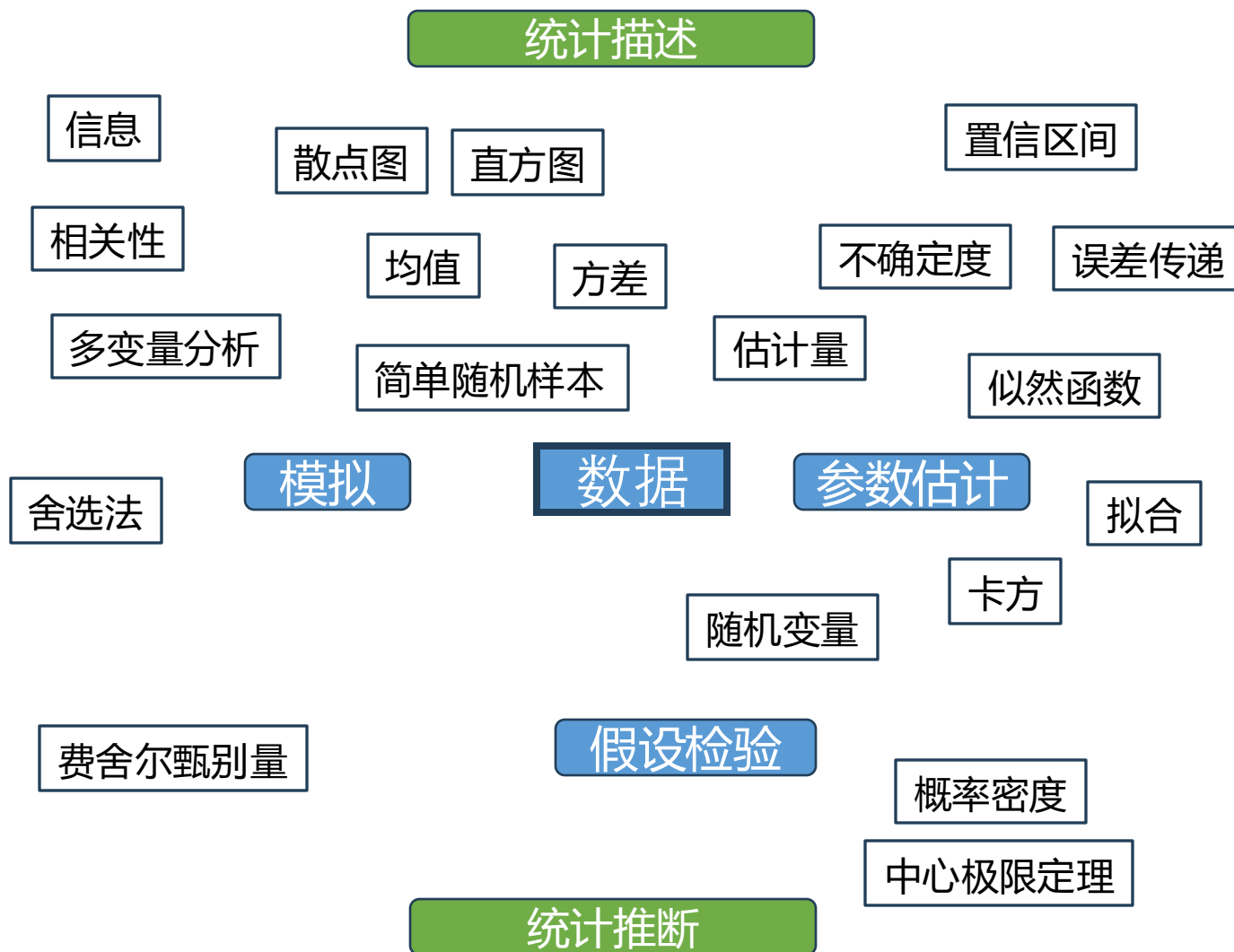
73 79 72 62 67 60 60 67 78 68 66 75 76 73 75 64 70 69 73 59 70 73 64 72 64 69  
69 71 69 71 77 69 72 71 67 72 63 66 68 76 71 76 68 71 63 65 65 66 73 73 73 67  
70 65 71 69 78 67 65 69 71 71 72 73 72 69 66 66 70 60 72 62 53 65 74 65 68 69  
67 75 64 76 72 76 78 67 67 67 69 79 71 67 71 68 71 65 66 65 78 76 71 70 67 65  
67 64 73 67 74 79 74 71 73 67 66 76 68 74 76 65 77 67 71 67 71 77 63 66 70 62  
68 74 67 67 67 77 65 68 79 72 71 77 68 70 73 67 81 70 74 71 79 62 67 63 68 76  
73 81 76 73 68 72 76 61 69 73 71 80 68 70 62 76 58 68 68 64 68 78 69 65 70 70  
64 75 73 72 60 86 68 68 64 60 68 71 70 75 70 67 69 67 73 65 66 71 70 70 73 66  
72 71 71 64 76 75 72 72 71 72 72 71 75 68 73 70 64 76 72 75 79 70 64 70 67 70  
75 70 83 69 61 70 66 69 71 72 70 76 73 62 71 60 73 74 70 68 68 70 78 71 69 71  
73 73 75 65 71 67 60 70 77 71 74 64 74 73 60 77 73 70 69 66 70 78 69 75 66 71  
75 75 74 69 74 70 75 77 75 66 72 68 72 61 75 65 69 68 65 73 82 67 75 67 80 71  
79 72 71 68 73 70 67 75 74 69 63 63 72 70 73 63 70 70 59 78 76 66 72 79 65 71  
76 72 69 69 73 70 77 73 83 66 68 67 69 73 76 65 71 70 71 65 78 71 67 70 72 75  
67 79 72 64 62 79 68 70 61 65 68 71 73 60 60 68 71 74 75 69 73 70 68 ...

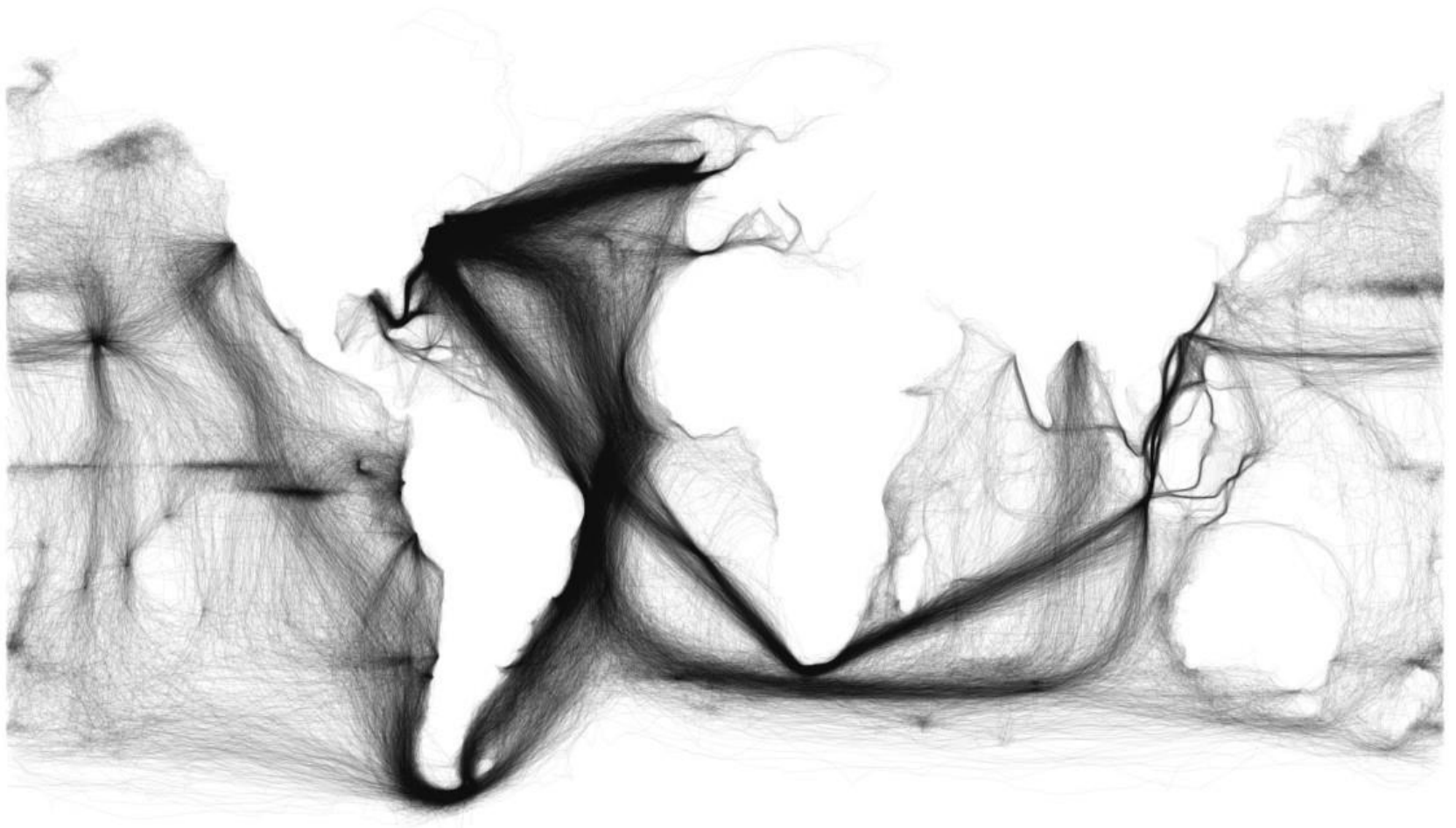


## ➤ 推断统计学

- 内容：参数估计、假设检验等
- 目的：对总体特征作出推断









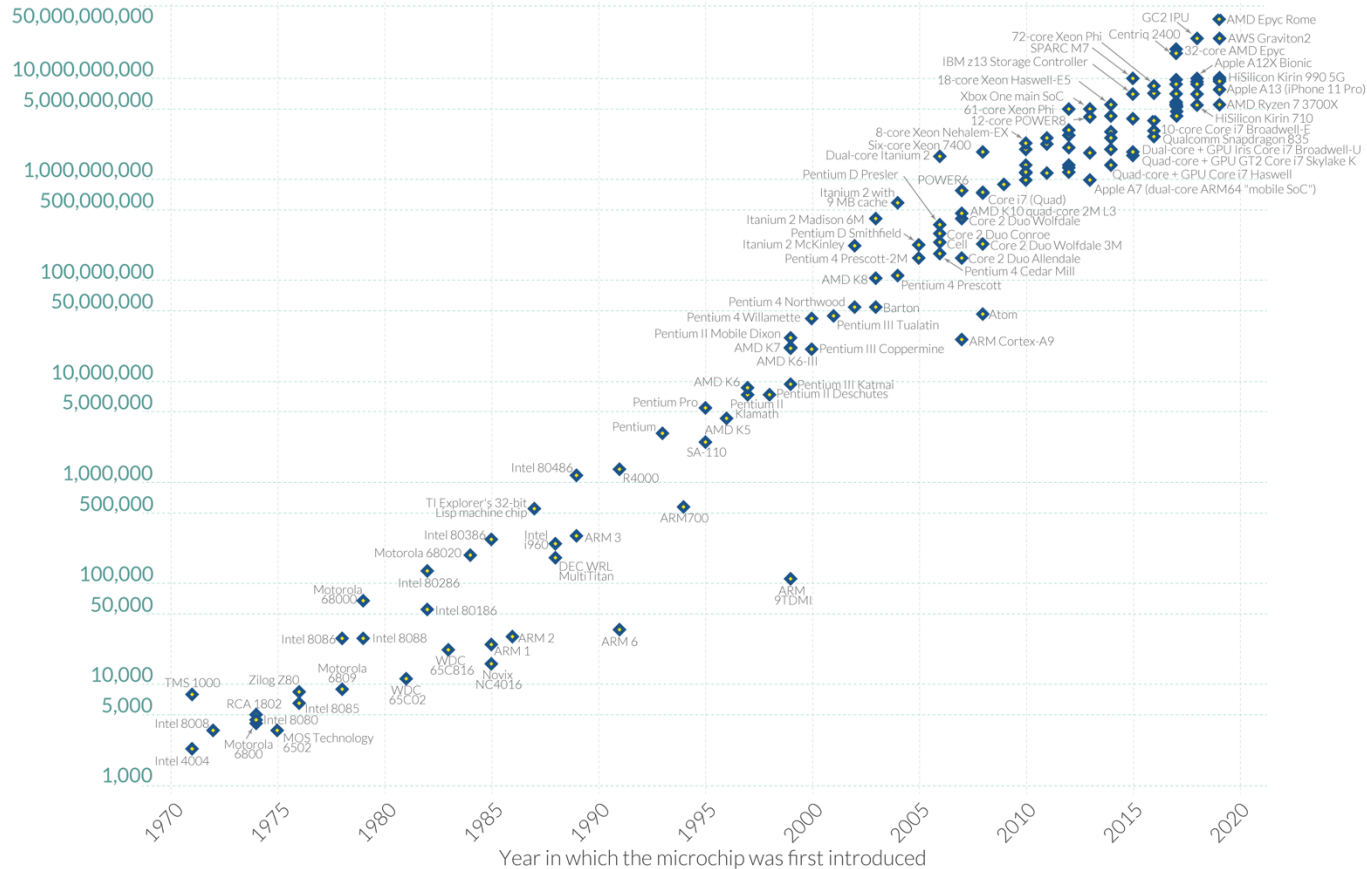
# 摩尔定律

Moore's Law: The number of transistors on microchips has doubled every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Our World  
in Data

## Transistor count

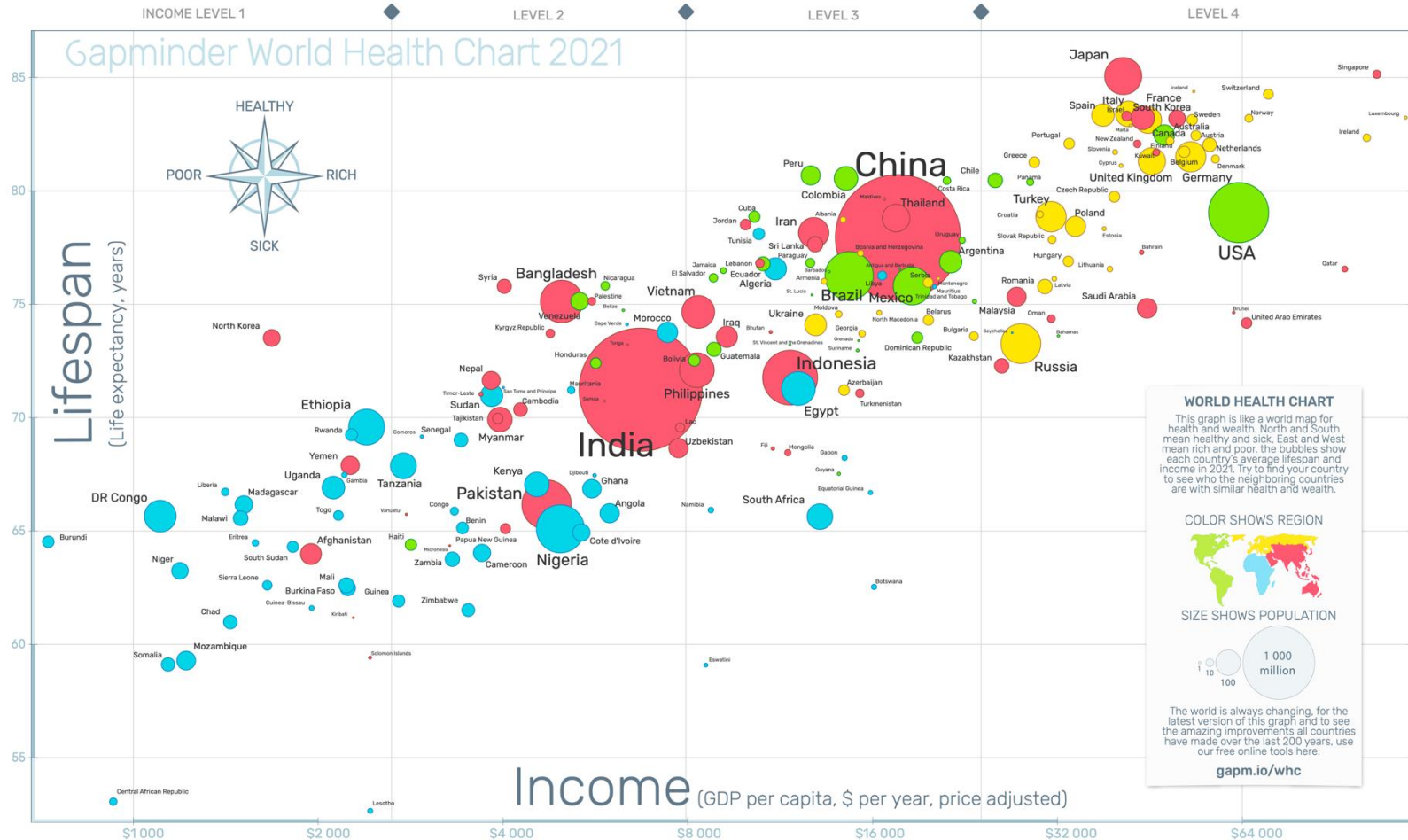


Data source: Wikipedia ([wikipedia.org/wiki/Transistor\\_count](https://wikipedia.org/wiki/Transistor_count))

OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.

# World Health Chart 2021



SOURCES - INCOME: World Bank's GDP per capita, PPP (2017 international \$) extended to 2021 with IMF's projections. X-axis uses log-scale to make a doubling income show the same distance on all levels.  
 POPULATION and LIFE EXPECTANCY: Data from UN Population Prospects 2019.  
 LICENSE: Our charts are freely available under Creative Commons Attribution License.  
 Please copy, share, modify, integrate, and even sell them, as long as you mention: "Based on a free chart from [www.gapminder.org](https://www.gapminder.org)".



VERSION 2022.1

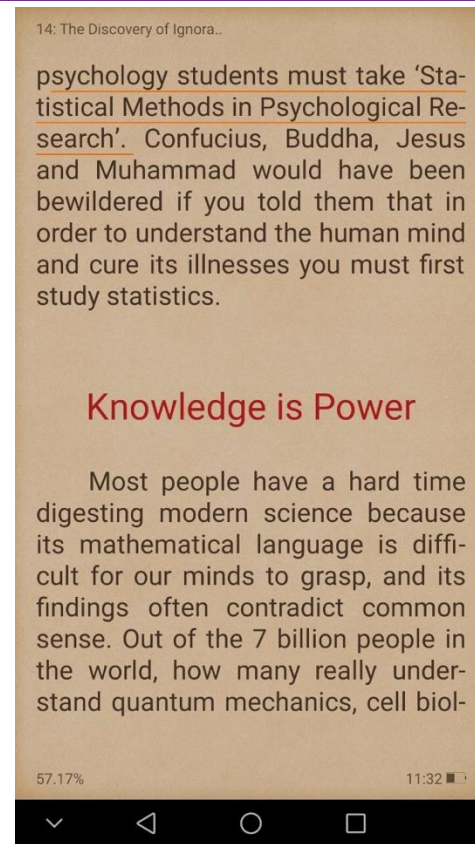
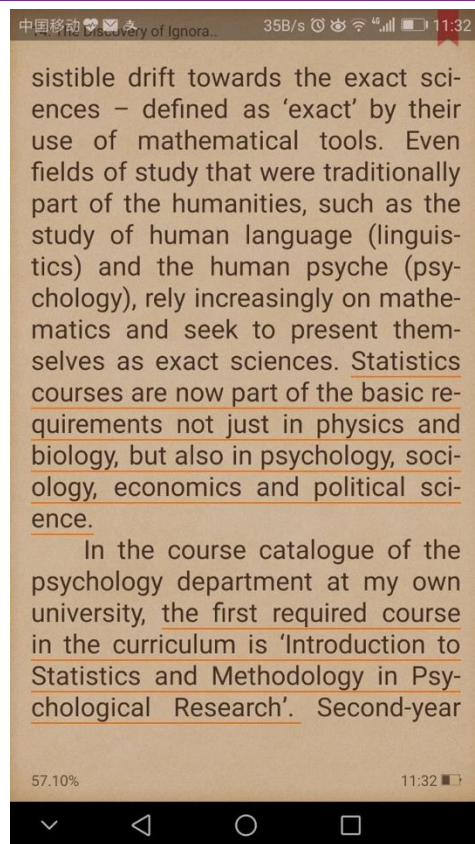
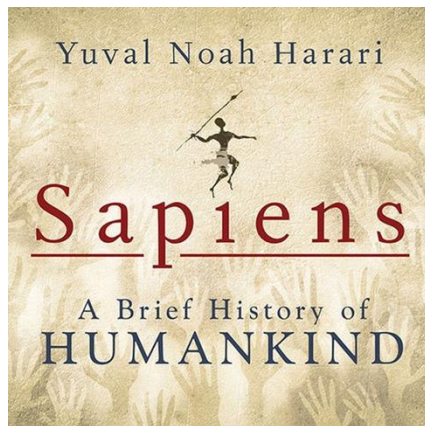
# 为什么统计学应用广泛？

---

- 统计学研究随机现象，随机现象无处不在
- 21世纪是什么的世纪？
  - 甲：21世纪是生物学的世纪
  - 乙：21世纪是计算机的世纪
  - 丙：21世纪是大数据的世纪
  - 丁：.....

如果说21世纪是大数据的世纪，  
那么21世纪实际上也是统计学的世纪。

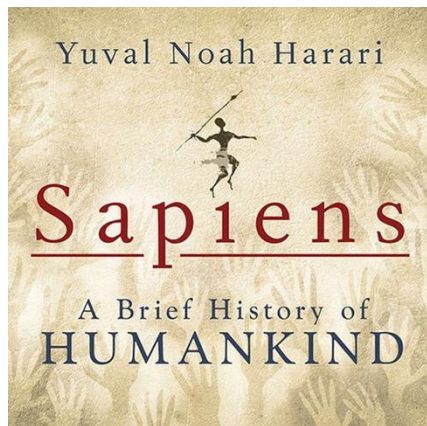
# 统计数据分析的重要性



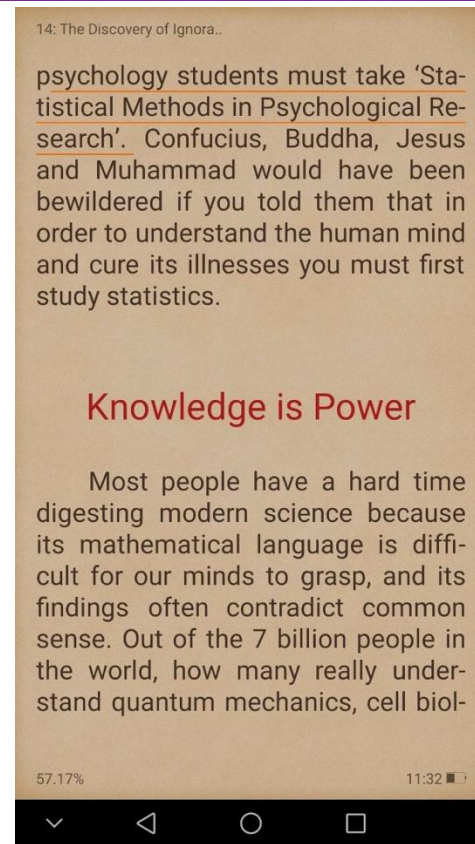
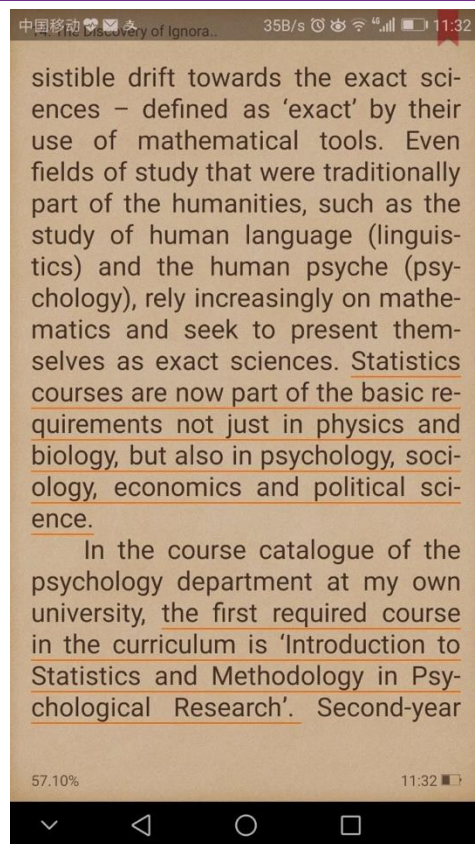
.....统计学如今不仅是物理学和生物学的基本要求之一，也是心理学、社会学、经济学和政治科学的基本要求。



# 统计数据分析的重要性（续）



.....第一门必修课是《心理学研究中的统计学和方法学导论》，心理学系二年级学生必须学《心理学研究中的统计方法》。



如果你告诉孔子、佛陀、耶稣和穆罕默德，为了理解人类的思想 and 治愈人类的疾病，你必须首先学习统计学，他们会感到困惑。

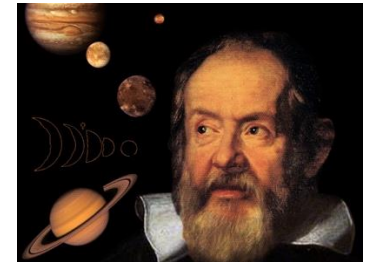
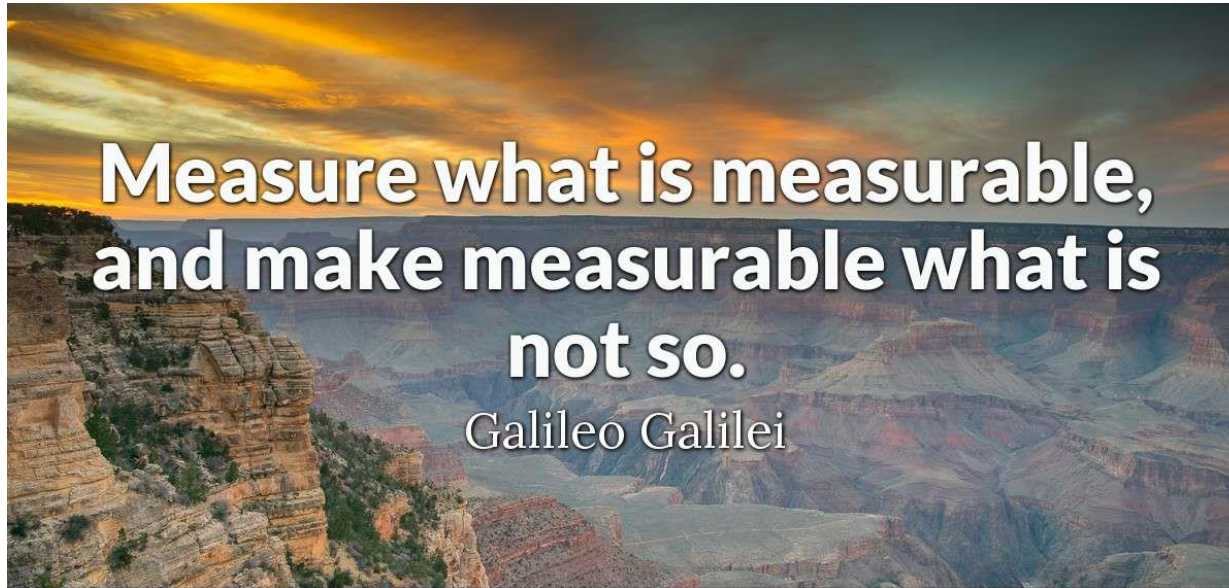
（此处致谢百度翻译）

---

统计学为什么对于物理学很重要？

# 伽利略的箴言

---



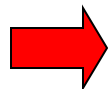
**“可测者，测之；不可测者，使之可测。”**

大伽：这确实是我说的

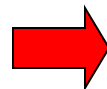
# 认知的过程

---

理论假设



实验设计



收集数据



数据分析



估计参数值与相应的误差范围，检验在何种程度上理论与实验数据相符。

问题：  
如何科学地给出物理结论？  
如何评价这种检验？  
如何不断改进认识？

贝叶斯定理对人类认知至关重要



# 为什么感觉不到统计学对物理很重要？

---

## ➤ 学习和研究的区别

- 万有引力跟统计学有关系吗？

$$F = \frac{GMm}{r^2}$$

- 欧姆定律跟统计学有关系吗？

$$V = IR$$

在**学习**科学时，你可能会发现统计学没什么用；  
但一旦你开始**研究**科学，并想知道测量的真正含义是什么，  
统计学就变得至关重要。

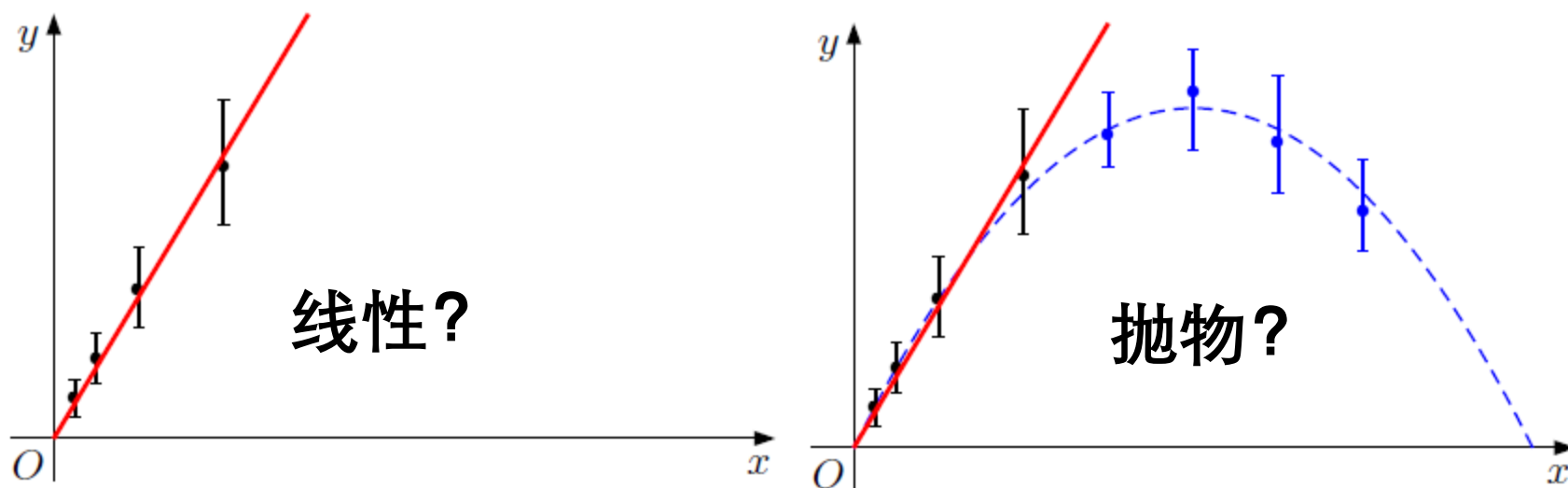
# 物理实验的目的是什么？

---

验证已有理论，或者揭示新规律

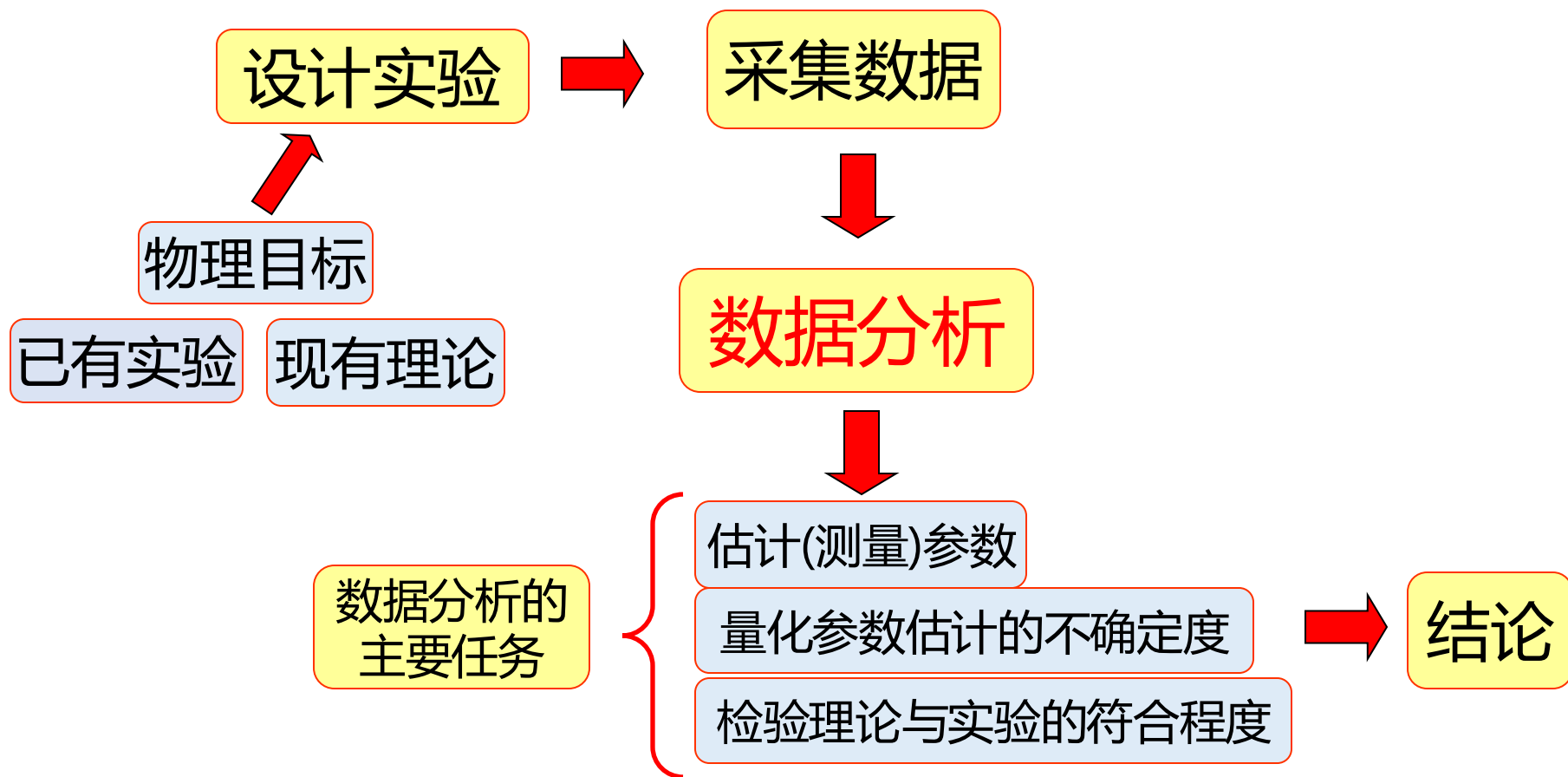
问题：如何实现？如何表述测量结果或发现？

# 拓展测量范围与提高测量精度的重要性



通过拓展测量范围（例如 $x$ 更大的区域），有可能揭示新的规律；  
通过提高测量精度，也可能揭示新的规律。

# 如何实现？如何科学地给出结论？



概率与统计在所有过程都起着重要作用

# 随机性的来源

---

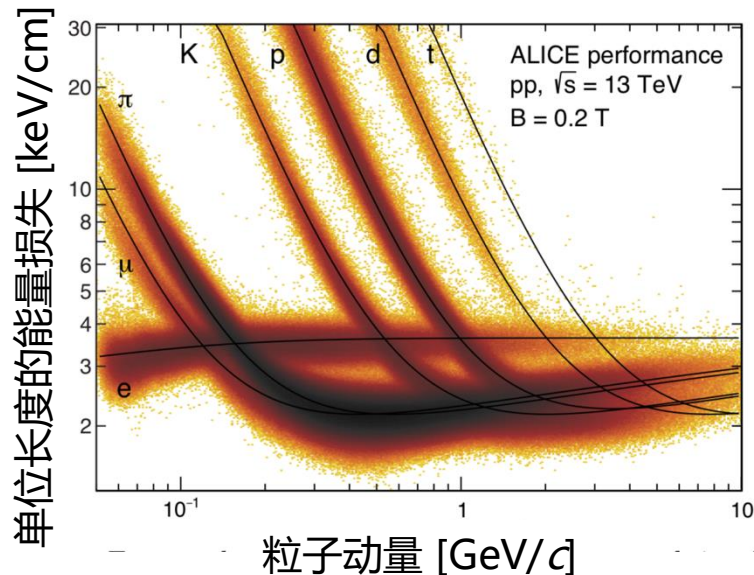
## ➤ 物理中随机性的主要来源

- 理论本身是非确定的
  - ✓ 量子力学 (...量力学)
- 测量的随机不确定性
  - ✓ 即使没有量子效应也存在
- 有些因素原则上可以确定但实际不确定
  - ✓ 受限于所需费用、时间等

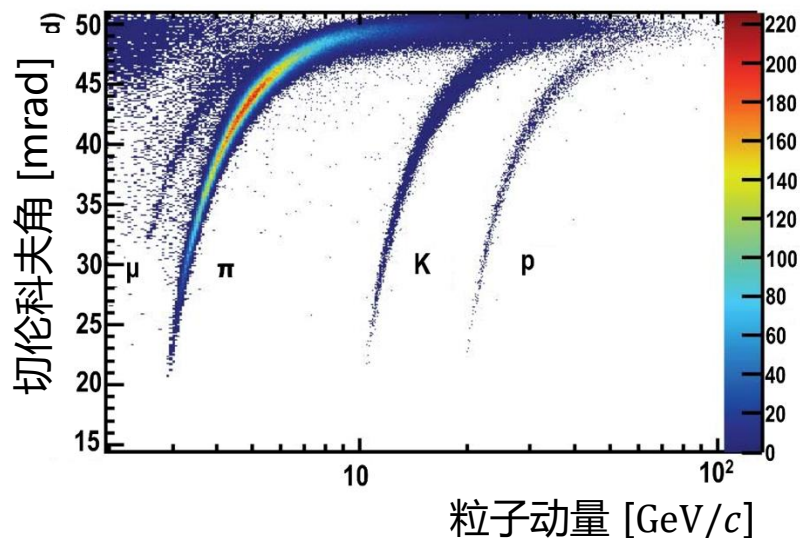
这些不确定性可以用概率量化描述

# 粒子鉴别

电离能损  $dE/dx$



切伦科夫效应



- 模式识别或假设检验
  - 指纹识别, 人脸识别, 核酸检测

# 概率与统计的差别

---



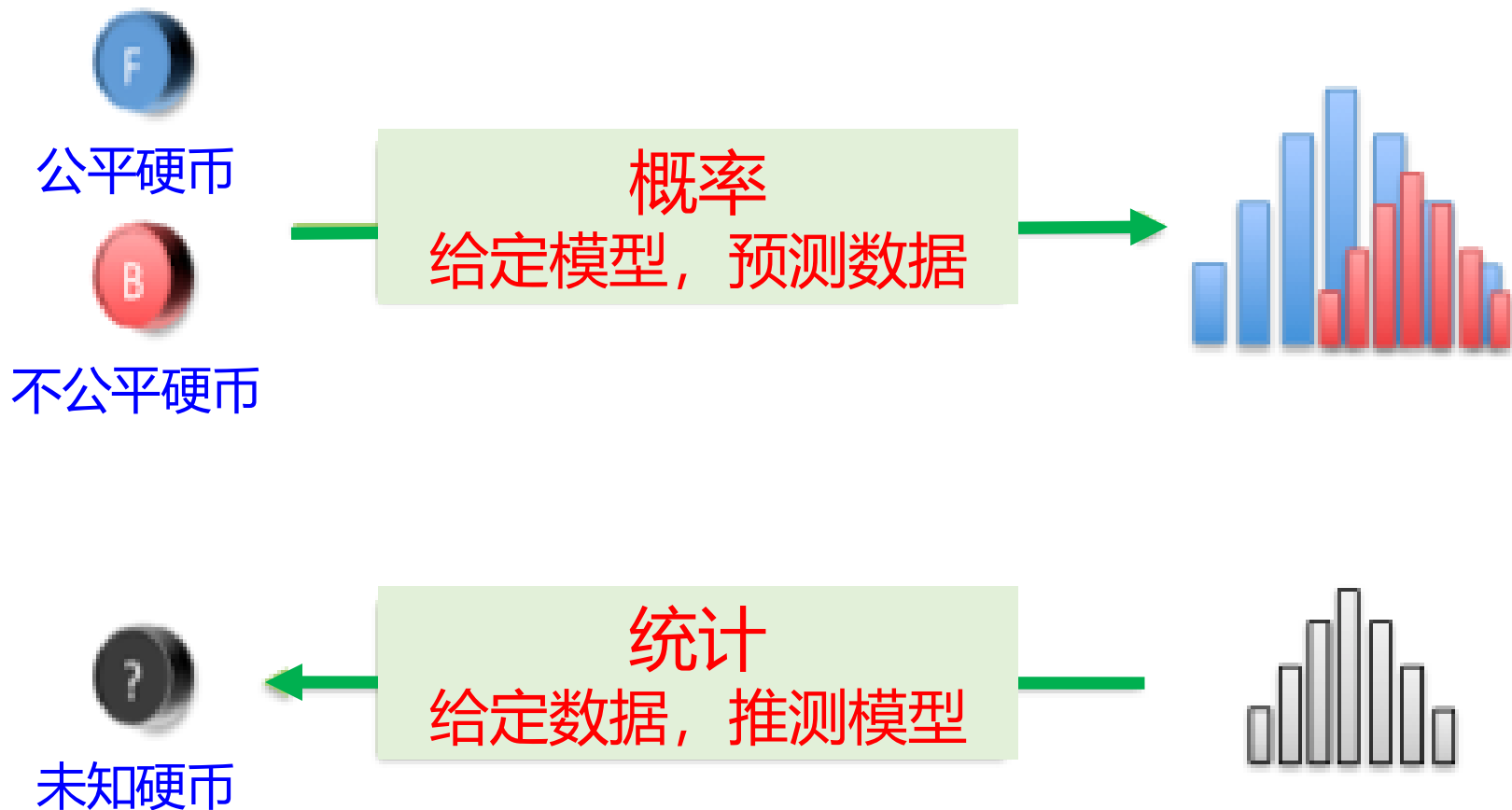
**概率：给定模型，预测数据。**

例如，已知桶里50白球，50黑球。  
问：手中的球什么颜色？

**统计：给定数据，预测模型。**

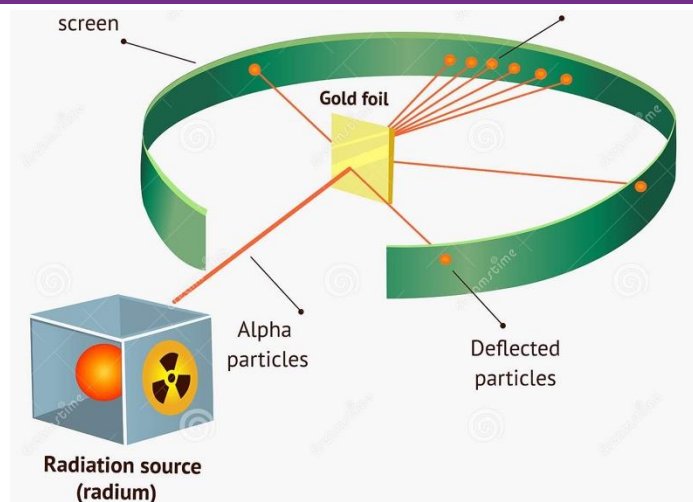
例如，已知手里4白球，6黑球。  
问：桶中白球黑球的比例是多少？

# 概率与统计的差别





# 实验观测



观察某一类型的  $n$  个事例

## 实验测量

给出每个事例的特征量（散射角，能动量 ...）。

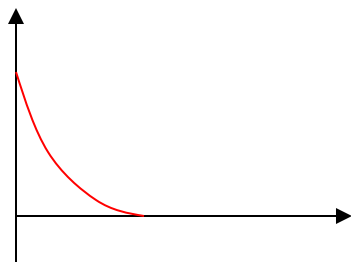
## 理论预言

给出上述各特征量的分布，而且可能还会包含自由参数。

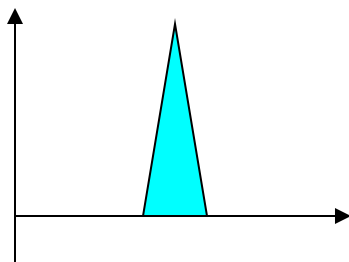
实验数据 → 统计分析 → 物理结论

# 数据背后的物理图像是什么？

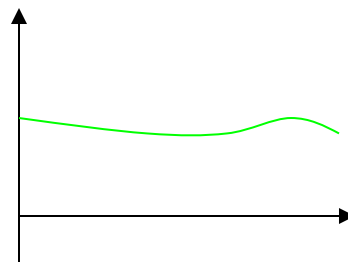
---



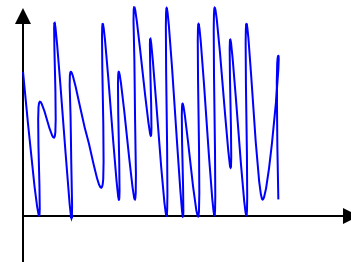
原初物理



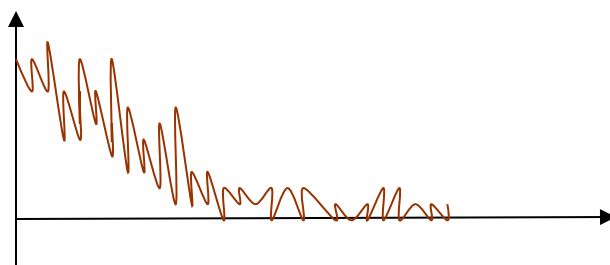
分辨率



探测效率



本底噪音



实验数据

# 统计学是工具、语言

---

- 对工具的熟练程度越高，能用它来做的事情就越多，也越简单



- 掌握一门“外语”很重要，正确地表达自己，也能“听懂”别人在说什么

# 如何表述结果？ 如何理解结果？

## Testing Real Quantum Theory in an Optical Quantum Network

Zheng-Da Li,<sup>1,2,\*</sup> Ya-Li Mao,<sup>1,2,\*</sup> Mirjam Weilenmann<sup>3</sup>, Armin Tavakoli,<sup>3,4</sup> Hu Chen,<sup>1,2</sup> Lixin Feng,<sup>1,2</sup> Sheng-Jun Yang,<sup>1,2</sup> Marc-Olivier Renou<sup>5</sup>, David Trillo<sup>6</sup>, Thinh P. Le,<sup>3</sup> Nicolas Gisin,<sup>6,7</sup> Antonio Acín,<sup>5,8</sup> Miguel Navascués,<sup>3</sup> Zizhu Wang (王子竹)<sup>9</sup> and Jingyun Fan<sup>1,2,†</sup>

...We experimentally demonstrate quantum correlations in a network of three parties and two independent EPR sources that violate the constraints of real quantum theory **by over 4.5 standard deviations**, hence disproving real quantum theory as a universal physical theory.

<sup>1</sup>Shenzhen Institute for Quantum Science and Engineering and Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China  
<sup>2</sup>Guangdong Provincial Key Laboratory of Quantum Science and Engineering, Shenzhen 518055, China  
<sup>3</sup>Institute for Quantum Optics and Quantum Information, University of Innsbruck, Innsbruck 6020, Austria  
<sup>4</sup>Institute for Quantum Optics and Quantum Information, University of Innsbruck, Innsbruck 6020, Austria  
<sup>5</sup>ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain  
<sup>6</sup>Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China

✉ (Received 27 October 2021; accepted 29 November 2021; published 24 January 2022)

Quantum theory is commonly formulated in complex Hilbert spaces. However, the question of whether complex numbers need to be given a fundamental role in the theory has been debated since its pioneering days. Recently it has been shown that tests in the spirit of a Bell inequality can reveal quantum predictions in entanglement swapping scenarios that cannot be modeled by the natural real-number analog of standard quantum theory. Here, we tailor such tests for implementation in state-of-the-art photonic systems. We experimentally demonstrate quantum correlations in a network of three parties and two independent EPR sources that violate the constraints of real quantum theory by over 4.5 standard deviations, hence disproving real quantum theory as a universal physical theory.

## Ruling Out Real-Valued Standard Formalism of Quantum Theory

Cheng Chen<sup>1,2,3,\*</sup> Can Wang,<sup>1,2,3,\*</sup> Feng-Ming Liu,<sup>1,2,3,\*</sup> Jian-Wen Wang,<sup>1,2,3</sup> Chong Ying,<sup>1,2,3</sup> Jing-Xia Shang,<sup>1,2,3</sup> Yulin Wu,<sup>1,2,3</sup> M. Gong,<sup>1,2,3</sup> H. Deng,<sup>1,2,3</sup> F.-T. Liang,<sup>1,2,3</sup> Qiang Zhang,<sup>1,2,3</sup> Zhi Peng,<sup>1,2,3</sup> Xiaobo Zhu<sup>1,2,3</sup>, Adán Cabello<sup>4,5</sup>, Chao-Yang Lu<sup>1,2,3</sup> and Jian-Wei Pan<sup>1,2,3</sup>  
<sup>1</sup>Heifei National Laboratory for Physical Sciences at Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui 230026, China  
<sup>2</sup>Centre for Excellence and Synergetic Innovation Centre in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai 201315, China  
<sup>3</sup>Shanghai Research Center for Quantum Sciences, Shanghai 201315, China  
<sup>4</sup>Departamento de Física Aplicada II, Universidad de Sevilla, E-41012 Sevilla, Spain  
<sup>5</sup>Instituto Carlos I de Física Teórica y Computacional, Universidad de Sevilla, E-41012 Sevilla, Spain

✉ (Received 30 November 2021; accepted 7 December 2021; published 24 January 2022)

Standard quantum theory was formulated with complex-valued Schrödinger equations, wave functions, operators, and Hilbert spaces. Previous work attempted to simulate quantum systems using only real numbers by exploiting an enlarged Hilbert space. A fundamental question arises: are the complex numbers really necessary in the standard formalism of quantum theory? We answer this question affirmatively by demonstrating that the real-number analog of quantum theory is not universal. We rule out the real-number analog, by revealing a contradiction between high-fidelity multiparty quantum experiment and the real-number analog. Our results disprove the real-number formulation and establish the indispensable role of complex numbers in the standard quantum theory.

...we faithfully realize the quantum game based on deterministic entanglement swapping with a state-of-the-art fidelity of **0.952**. Our experimental results violate the real-number bound of 7.66 by **43 standard deviations**.

# 统计学术语

---

- 样本，总体
- 估计，估计量
- 参数估计，区间估计
- 假设检验，显著性水平
- 置信水平，置信区间
- 弃真错误，取伪错误
- .....

# 学习概率统计-重在应用

---

- 统计方法是从大量（实验）数据中总结提取一般规律的必要手段，概率论是统计方法的基础。
- 桥接数学与应用科学的数学工具
- 几乎所有与数据有关的学科的语言

在保持一定的数学严谨性的前提下，  
重点强调统计方法在物理学科中的应用。

例如，

如何根据物理目标完成实验的设计？

如何从实验数据中提取物理规律？如何规范表述实验结果？

# 课程目的

---

实验设计 → 实验数据 → 统计分析 → 物理结论

➤ 介绍实验物理数据分析的基本数学工具

- 概率与统计理论

- ✓ 假设检验
- ✓ 参数估计, 区间估计
- ✓ 系统不确定度的分析

➤ 通过实践熟悉统计数据分析的具体应用

- 数据处理工具

- ✓ 直方图、散点图、等高线图等绘图技能
- ✓ 参数拟合、假设检验
- ✓ 蒙特卡罗模拟





---

# Backup

# 基础物理实验课的一道习题

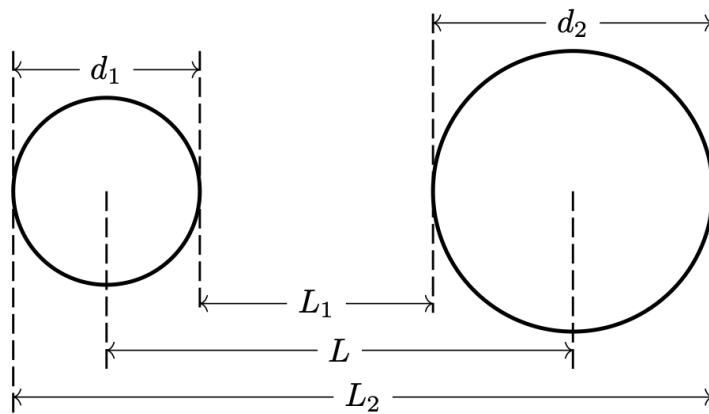
## 实验七 “测量误差与数据处理” 第4题

现有一个箱体，其内有两个转子，要测量两转子的轴心距离  $L$ ，如图所示。现在测量手段所能达到的精密度分别为

$$\begin{aligned}\sigma_{d_1} &= 0.5 \mu\text{m}, & \sigma_{d_2} &= 0.7 \mu\text{m}, \\ \sigma_{L_1} &= 0.8 \mu\text{m}, & \sigma_{L_2} &= 1.0 \mu\text{m}.\end{aligned}$$

试问怎么安排测量能使  $\sigma_L$  最小？

(假设对  $L_1$ 、 $L_2$ 、 $d_1$  和  $d_2$  的测量均相互独立。)



# 用统计学语言描述：

---

已知两次观测结果，且两次观测结果不相互独立，

问题：如何构造一个估计量，从而给出待观测量的最佳估计？

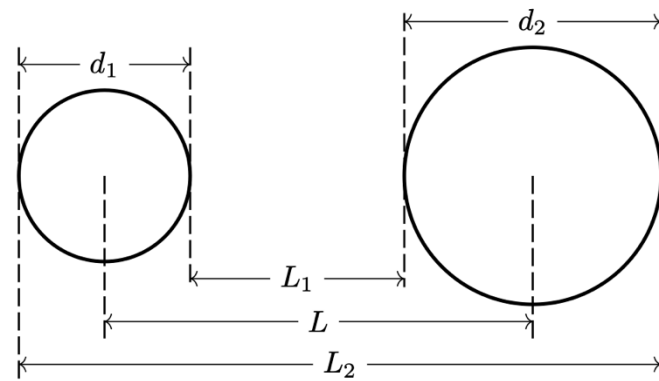
# 一种典型的求解

已知:  $\sigma_{d_1} = 0.5 \mu\text{m}$ ,  $\sigma_{d_2} = 0.7 \mu\text{m}$ ,  
 $\sigma_{L_1} = 0.8 \mu\text{m}$ ,  $\sigma_{L_2} = 1.0 \mu\text{m}$ 。

$L$  可以表示为

$$L = L_1 + (d_1 + d_2)/2$$

$$L = L_2 - (d_1 + d_2)/2$$



$$L = L_1 + \frac{d_1 + d_2}{2} \Rightarrow \sigma_L = \sqrt{\sigma_{L_1}^2 + (\sigma_{d_1}^2 + \sigma_{d_2}^2)/4} = 0.91 \mu\text{m}$$

$$L = L_2 - \frac{d_1 + d_2}{2} \Rightarrow \sigma_L = \sqrt{\sigma_{L_2}^2 + (\sigma_{d_1}^2 + \sigma_{d_2}^2)/4} = 1.09 \mu\text{m}$$

$$L \text{ 也可表示为 } L = \frac{L_1 + L_2}{2} \Rightarrow \sigma_L = \sqrt{(\sigma_{L_1}^2 + \sigma_{L_2}^2)/4} = 0.64 \mu\text{m}$$

# 一种正确的求解方法

---

这实际上是统计学中构造有效估计量的问题

$$\hat{L} = \lambda \left[ L_1 + \frac{d_1 + d_2}{2} \right] + (1 - \lambda) \left[ L_2 - \frac{d_1 + d_2}{2} \right]$$

我们希望求出待定参数 $\lambda$ ，使得 $\sigma_{\hat{L}}$ 或 $\sigma_{\hat{L}}^2$ 最小。

$$\sigma_{\hat{L}}^2 = \lambda^2 \sigma_{L_1}^2 + (1 - \lambda)^2 \sigma_{L_2}^2 + \left( \lambda - \frac{1}{2} \right)^2 (\sigma_{d_1}^2 + \sigma_{d_2}^2)$$

$$\frac{d(\sigma_{\hat{L}}^2)}{d\lambda} = 2\lambda \sigma_{L_1}^2 + 2(1 - \lambda) \sigma_{L_2}^2 + 2 \left( \lambda - \frac{1}{2} \right) (\sigma_{d_1}^2 + \sigma_{d_2}^2)$$

$$\text{令 } \frac{d(\sigma_{\hat{L}}^2)}{d\lambda} = 0 \Rightarrow \lambda = \frac{2\sigma_{L_2}^2 + \sigma_{d_1}^2 + \sigma_{d_2}^2}{2(\sigma_{L_1}^2 + \sigma_{L_2}^2 + \sigma_{d_1}^2 + \sigma_{d_2}^2)}$$

$$\Rightarrow \sigma_{\hat{L}} = 0.63 \mu\text{m}$$

# 另一种正确的求解方法

---

问题相当于用不同的方法对同一个量 $L$ 进行了两次测量：

$$y_1 = L_1 + \frac{d_1 + d_2}{2}, \quad \sigma_{y_1} = 0.91 \mu\text{m}$$

$$y_2 = L_2 - \frac{d_1 + d_2}{2}, \quad \sigma_{y_2} = 1.09 \mu\text{m}$$

我们希望合并两次测量结果。

如何合并？需要考虑什么问题？

统计学方法告诉我们如何处理

$$\hat{L} = \frac{y_1 + y_2}{2} \Rightarrow \sigma_{\hat{L}} = \frac{1}{2} \sqrt{\sigma_{y_1}^2 + \sigma_{y_2}^2} = 0.7 \mu\text{m?} \quad \text{对吗?}$$

# 问题讨论

---

- “明天降水概率是60% ”。这样的天气预报到底什么意思？
- 某实验测量电子质量，结果为 $0.511 \pm 0.001 \text{ MeV}/c^2$ 。这个表述具体是什么含义？
- 如果有人说，某一枚硬币不是公平的硬币，正面和反面的概率分别为51%和49%。  
如何设计一个实验验证这个说法对还是不对？
- 如果让你写个程序，利用摄像头信息判断进入教室的人的性别。你会如何实现？

