

Application of Statistics in Psychology

01 - Introduction

Hu Chuan-Peng

2023-02-22



教师介绍



- 2005.09 – 2009.06 湖北京大学法学学士；心理学学士（双学位）
- 2009.09 – 2012.06 湖北京大学发展与教育心理学硕士
- 2012.08 – 2017.06 清华大学心理学系，博士
- 2017.10 – 2020.10 德国莱布尼茨心理韧性研究所，博士后 Chinese Open Science Network (公众号：Open Science) 的发起人 研究兴趣：元科学 (meta-science) 、贝叶斯模型 (Bayesian modelling) 和自我认知 (self-cognition) 等

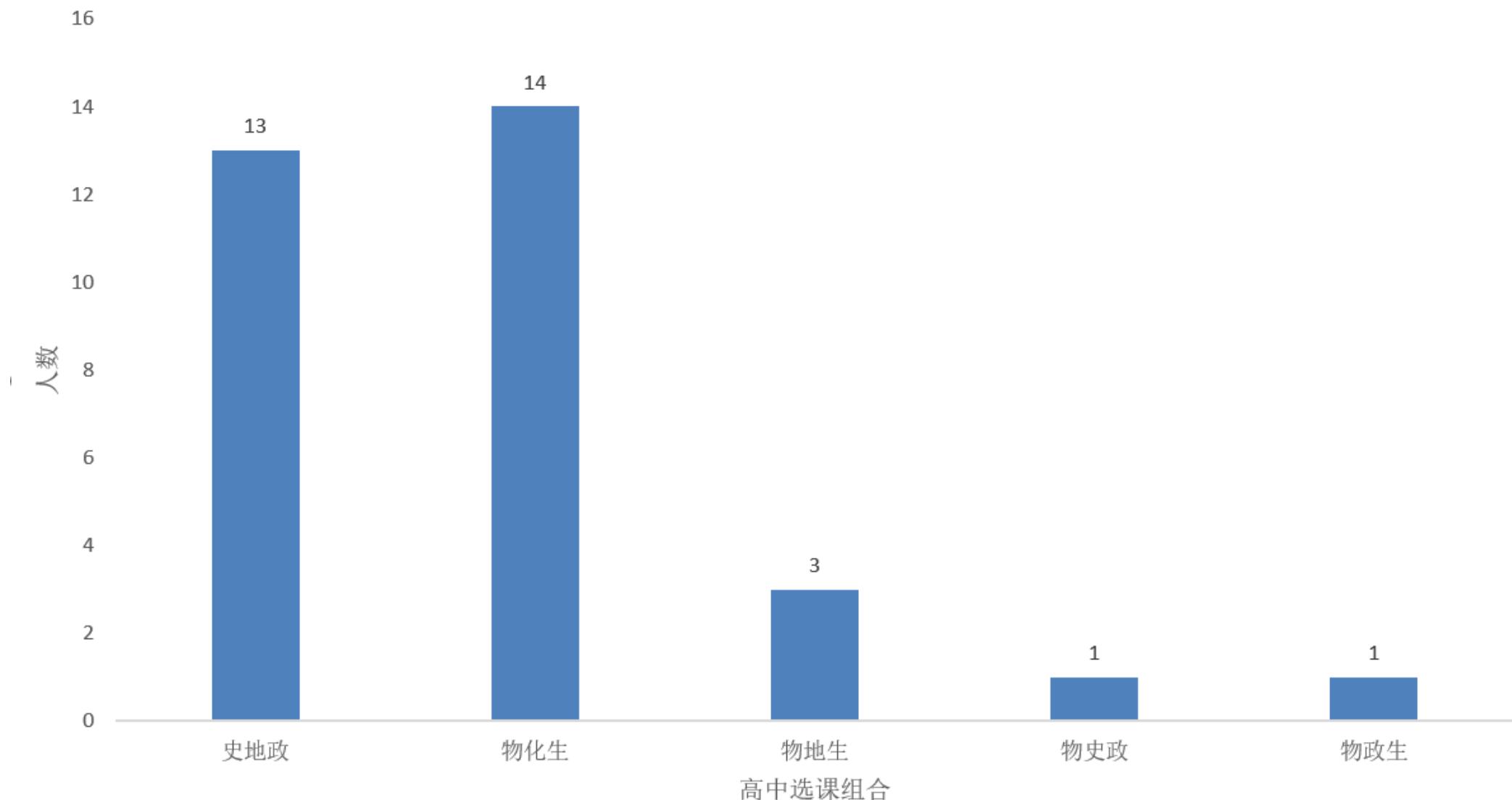
主页：

<https://xlyy.njnu.edu.cn/info/1073/8311.html>



学科基础

学科基础



职业规划



- 做一个自由度高一点的心理咨询师
- 如果可能，希望以后走科研道路
- 对心理学的分类，能从业的方向都还不是很清楚，目前打算考研吧
- 目前还是想先考研，职业的话主要是想在南京当心理老师跟考公，去公司做人事也在考虑方案中
- 先去我感兴趣的和心理学直接有关的公司（如KY，暂停实验室，白岛岩心）面试实习，把实习当作探索职业兴趣和锻炼、积累能力的过程和机会。尝试后再判断到底愿不愿意去做



你觉得这门课会教哪些内容？



- 数理统计相关知识，辅助心理学研究
- 心理学的研究方法、SPSS应用、数据分析、相关分析、回归分析
- 会教授统计学的一些原理和分析方法，还有与专业特色相切合的一些统计学知识，比如心理问卷的设计
- 研究在心理实验或调查中如何收集、整理、分析数字资料，以及如何根据这些资料所传递的信息作出科学推论
- 没有提前了解过，但是我觉得应该是有关心理测评的，能教我们如何通过统计的思维去大概分析、判断对方的心理状态



文、理數學居

111

第十一章

心理统计

卷之三

100

100

设计向导

学会分析数据

卷之三

了解卷土的應用

培养思维逻辑能力

第十一章

卷之三

嘉慶十一年

学习理论知识

数据的收集与运用

研究

第10章

三

分析技术

卷之三

三

统计方法

你希望在这门课上学到哪些内容？



对统计学的认识

更理性

完成心理学调查

提高数学逻辑能力

大致总结规律

处理大数据
理性决策

专业科学

提高数据处理的能力
复杂数据
提高数学逻辑分析能力

帮助规划

掌握实验研究
提供理论基础
提供技术支持
考研上岸 重难点把控
培养判断统计能力

运用数据分析



课程概况



心理统计学的两种视角

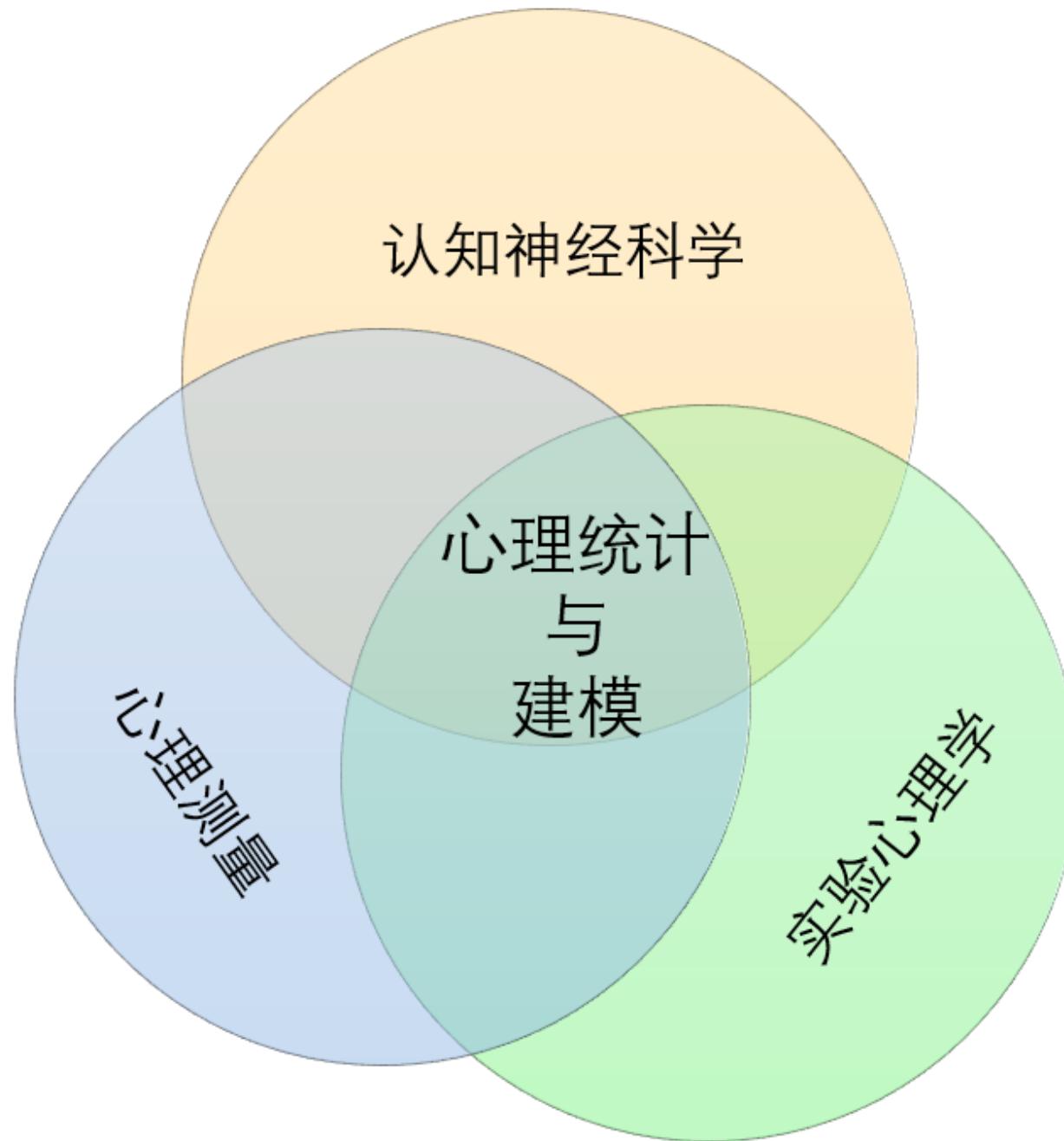
统计学

- 在心理学中的应用
- 在经济学中的应用
- 在生物学中的应用
- ...

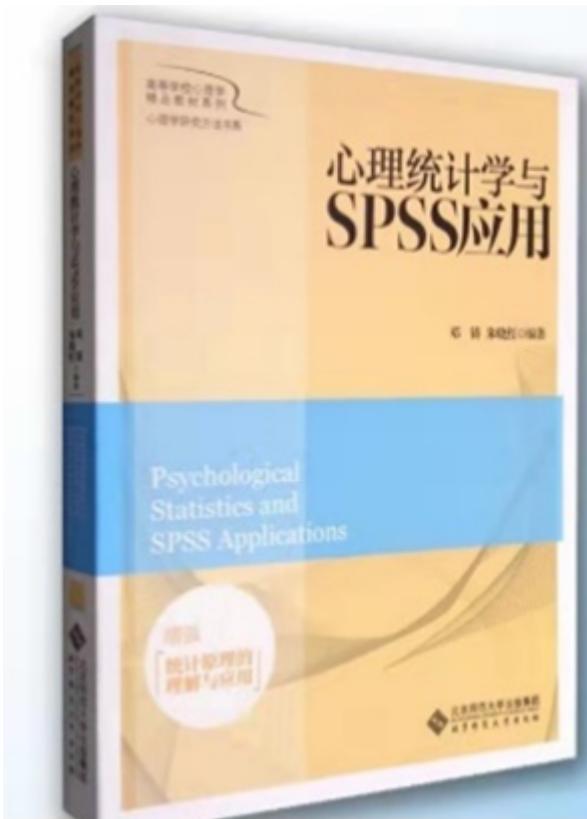
科学心理学(研究方法)

- 心理统计学
- 实验心理学/实验设计
- 心理测量
- 神经成像方法
- ...

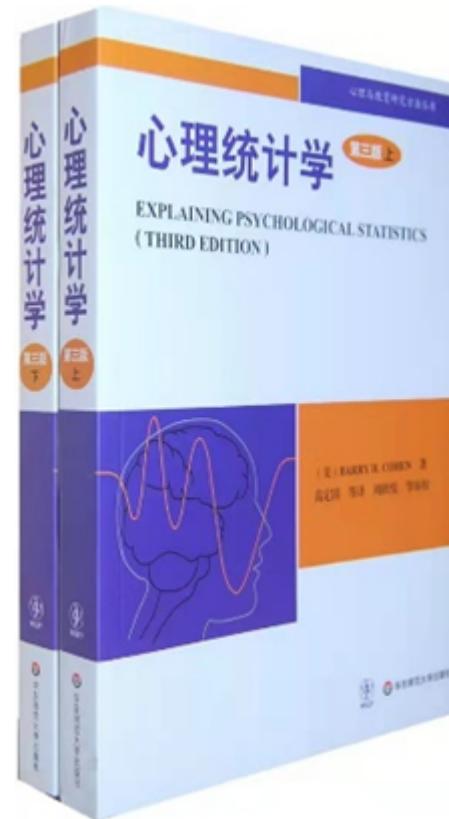




参考教材



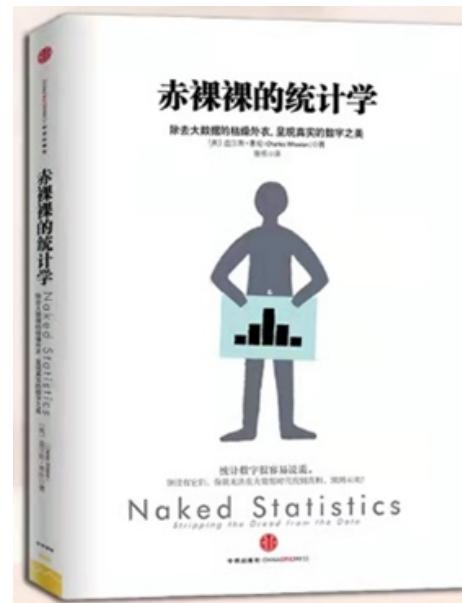
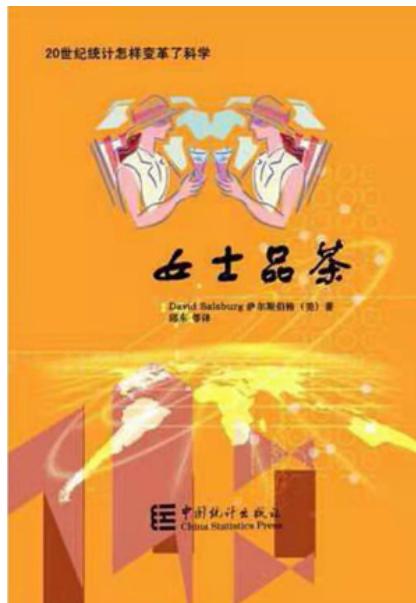
《心理统计学与SPSS应用》（第1版），邓铸等编著，北京师范大学出版社，2017年



《心理统计学》（第3版），(美) B.H.科恩编著，高定国、周欣悦译，华东师范大学出版社，2011年



课外读物



教学目标

1. Know it [as an educated citizen]

掌握基本的统计知识。

- 区分不同，比如分辨不同数据类型，数据的分布类型。
- 了解最常用的统计方法软件。
- 能够区分科普读物中正确与错误的统计推断。



2. Do it [as a trad. psychologist]

- 掌握用不同统计方法，以及统计软件的实现。
- 具体问题，具体分析，数据类型与分析方法相对应。
- 用软件实现需要的分析功能。



3. Do it appropriately [as a well-trained researcher]

基于workflow研究问题

- 可视化。
- 提炼核心问题(If p-value is the answer, what is the questions?)
- 方法与问题匹配(Appropriate model/technique)
- 可重复(Reproducible).



职业规划



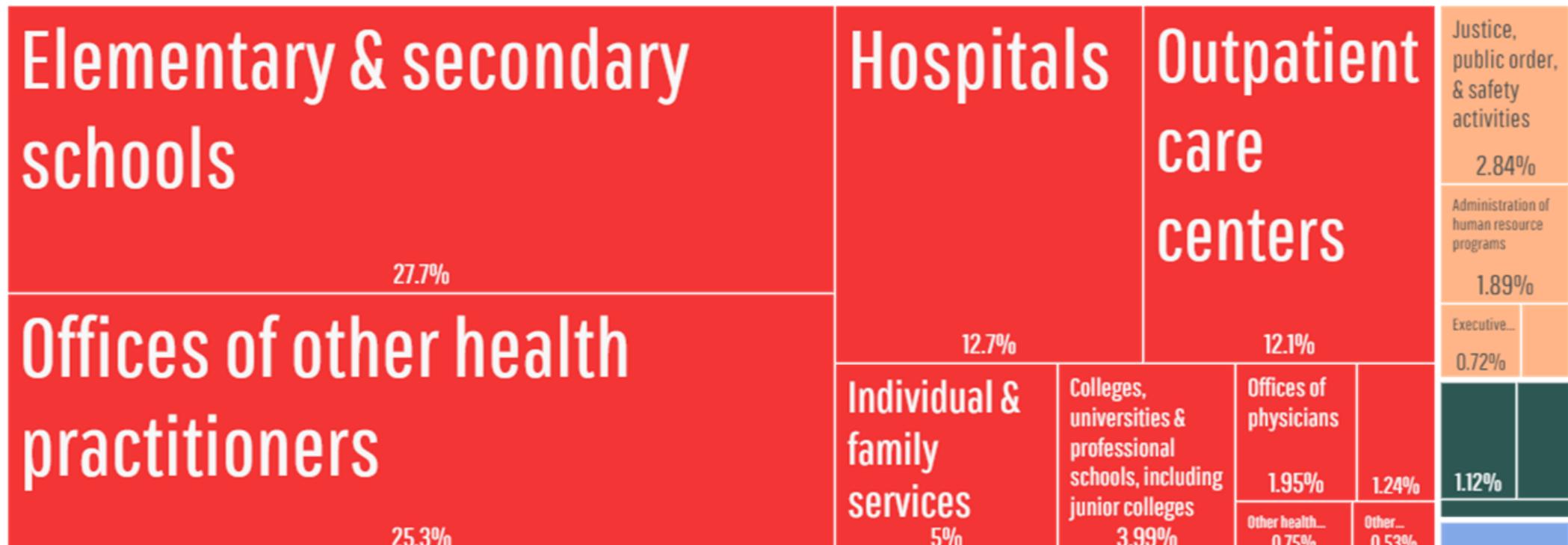
教学目标与职业规划的关系

- Know it. 拥有一些统计学常识，辨别心理健康数据，企业项目数据。
- Do it. 利用统计软件对数据进行处理，验证或发现规律。
- Do it correctly. 看懂数据，正确决策，靠谱的科普。
- Invent it. 在实践过程中，开发新的软件，做出统计方法的改进与改良。



心理学就业与心理统计学

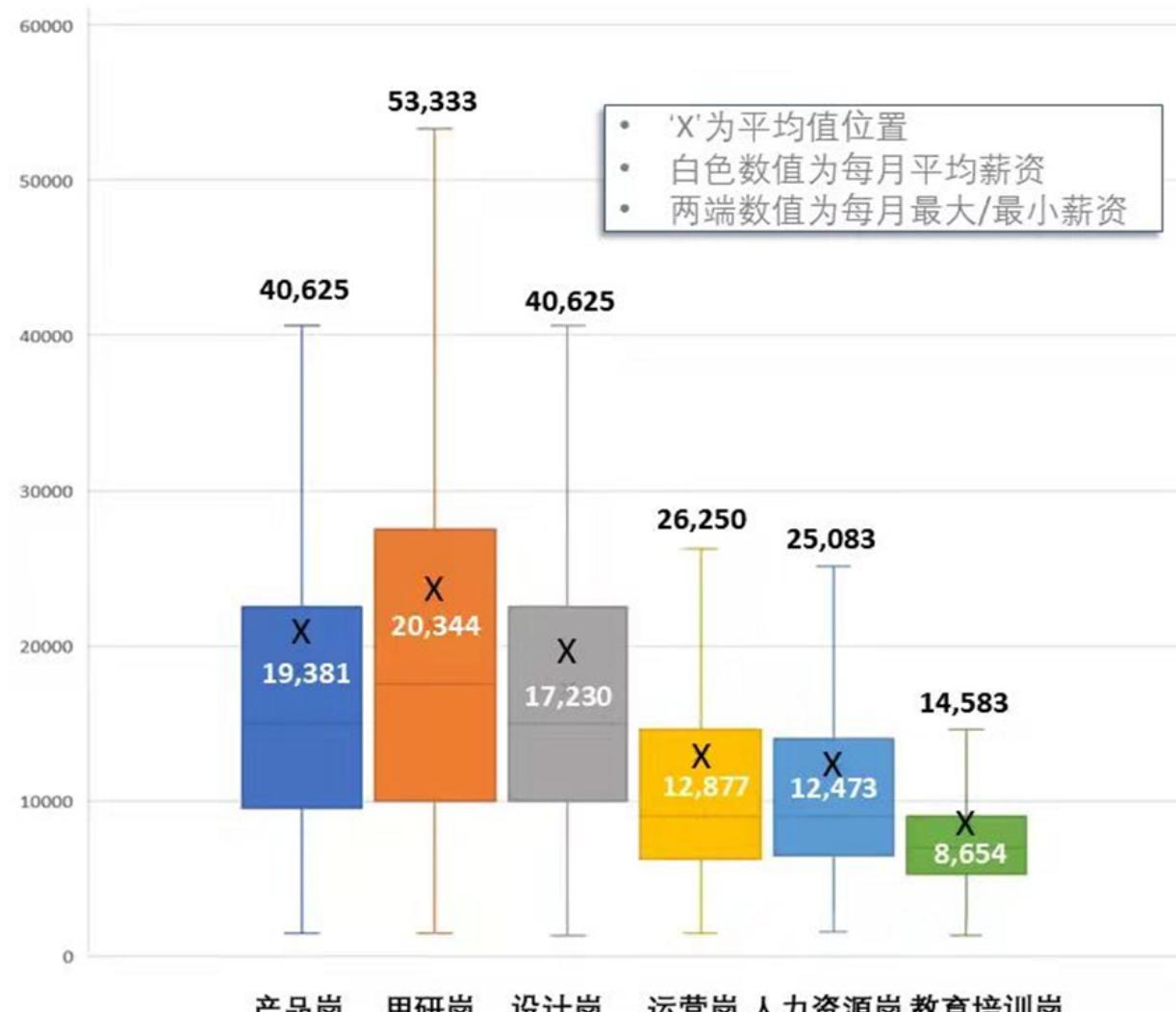
Total: 165,363



(2017年美国心理学就业方向，图片来源：DATA USA:Psychologists)



心理学就业与心理统计学



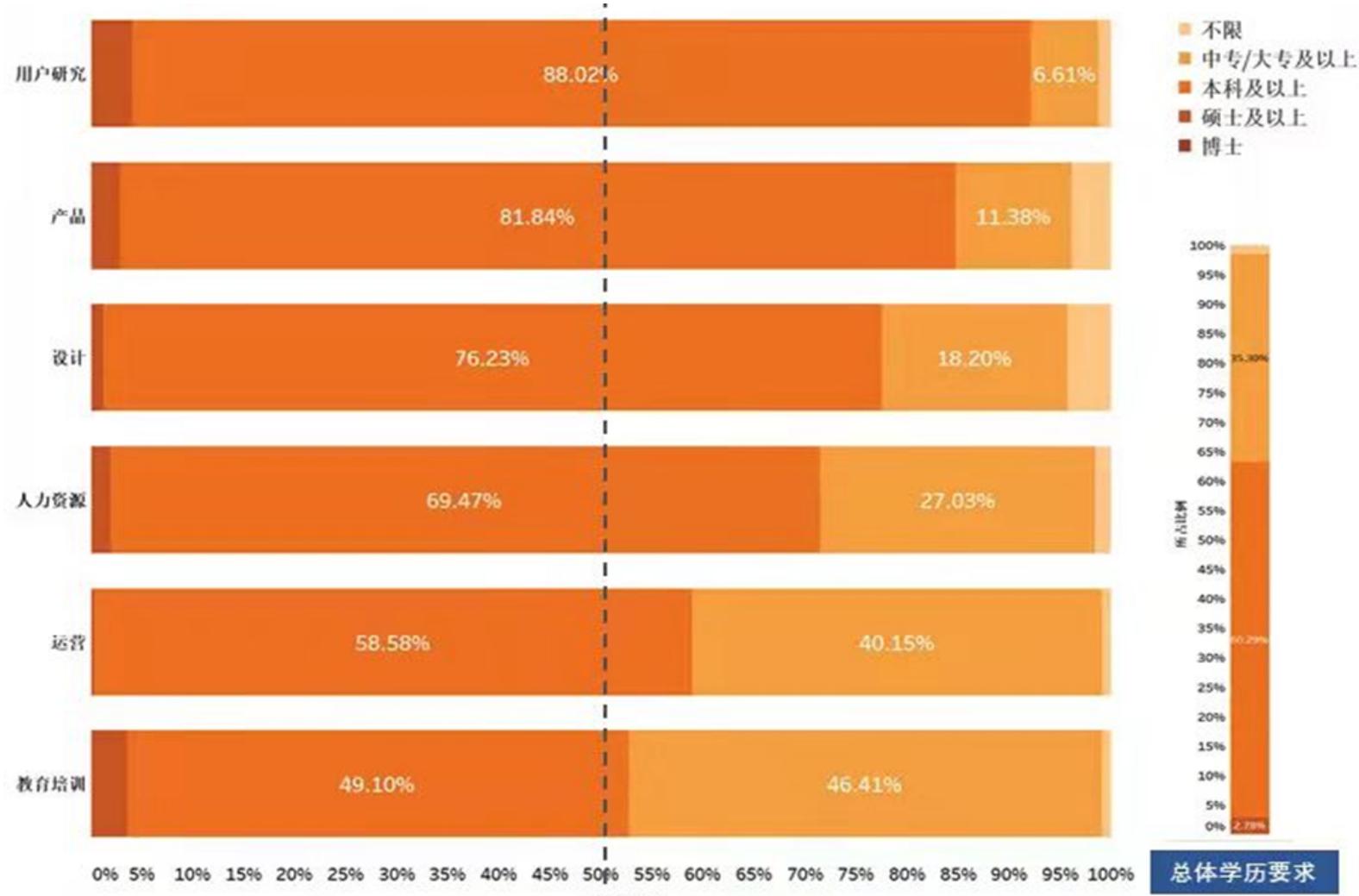
统计日期:2020年3月31日-4月6日, 利用python爬取

数据来源: 拉勾网、猎聘网、前程无忧三家招聘网站以“心理学”为搜索词的搜索结果

(中国心理学各岗位工资水平, 图片来源: 北师大赛客双创中心公众号)



心理学就业与心理统计学



统计日期:2020年3月31日-4月6日，利用python爬取

数据来源：拉勾网、智联网、前程无忧三家招聘网站以“心理学”为搜索词的搜索结果

(图片来源：北师大赛客双创中心公众号)



统计学与数字时代

2020 This Is What Happens In An Internet Minute



(图片来源：Twitter @OfficiallyChadd)

一分钟产生的数据量

数据素养不是数学技能而是生活技能



本课目标与就业及数字化社会

培养统计素养的理由：

个体：

- 数据消费者，避坑；
- 职业发展：潜在的（心理与行为）数据科学从业者。

组织：

- 数据是其最有价值的资产之一。
- 相同的数据可以在不同的应用程序和场景中使用和复用。
- 数据的正确使用为企业赋能，提高竞争力，如Tik Tok，京东。



市场：

- 拥有数据素养的雇员能够使企业在市场竞争中获益。
- 数字企业科技发展更快，数字化经济渗透，更好发展。

国家政策：

- 数字技术的使用能够促进传统企业转型升级。
- 数字技术能够驱动生活方式与治理方式变革。



本课目标与就业及数字化社会

数字经济的“四化框架”

生产要素

数据价值化



采集汇聚



确权定价



交易市场



安全保护

技术 资本
劳动 土地...

生产力

数字产业化



基础电信



软件及服务



电子信息制造



互联网

产业数字化



数字技术
农业边际
贡献

农业



数字技术
工业边际
贡献

工业



数字技术
服务业边际
贡献

服务

生产关系

数字化治理



多主体参与



技管结合



数字化
公共服务

(图片来源：中国信息通信研究院)



教学内容

序号	内容	序号	内容
1	统计学的历史与意义	9	参数估计2
2	概率与概率分布	10	假设检验(Fisher)
3	数据与数据可视化	11	P值
4	集中量数和离散量数	12	两类错误
5	离散变量概率分布	13	相关和相关显著性
6	正态分布	14	z检验和单样本t检验
7	标准分数	15	双样本t检验
8	参数估计1	16	线性代数基础



R



R Programming



考查方式

出勤(10%)

课堂测试成绩($10 \times 3 = 30\%$)

- 3次小测验

小组作业(20%)

- 自由分组
- 完整的workflow

期末考试(40%)



如何学好这门课

- 1.本课内容没有想象中那么困难，仅仅高中水平的数学知识。
- 2.理解 > 推导，方法与问题相匹配是关键。
- 3.编码/软件操作 > 手动运算，合理使用现代工作。
- 4.一图胜千言，理解和使用可视化工具。
- 5.广泛阅读课外科普书籍，可能会有意外收获。

Warnings: 学好 ≠ 考好, 理解 ≠ 会用, “做中学”



Any question?



统计学简史



"统计(statistics)"



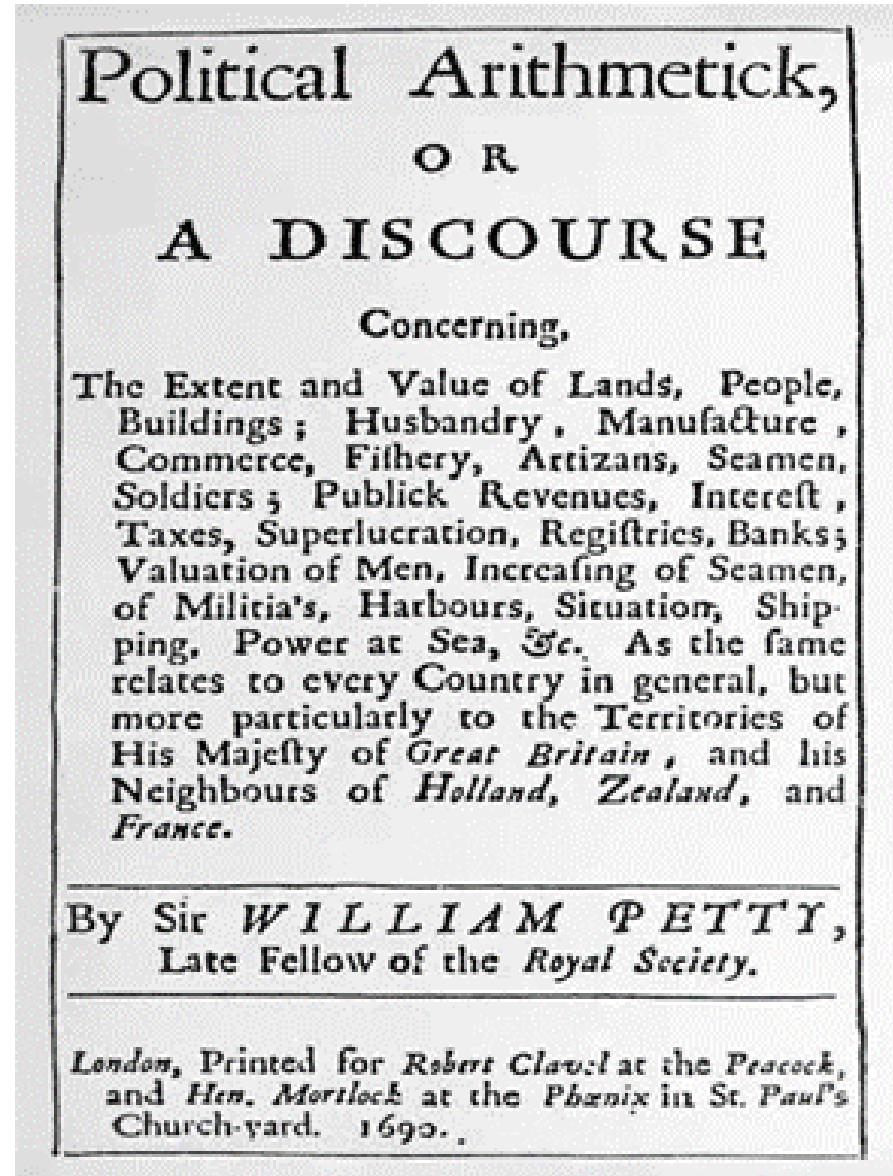
亚里士多德

国情调查(State of the Nation Survey):

- “统计(statistic)”来源于拉丁语“statisticum collegium”，意为国会。
- 亚里士多德(Aristotle)写了150多部“城邦编年史(city-state chronicles)”，其中包括对各个城邦在历史、管理、资源和财富方面的社会和经济状况进行比较分析。



"统计(statistics)"



政治算术



"统计(statistics)"

政治算术(Political Arithmetic)：

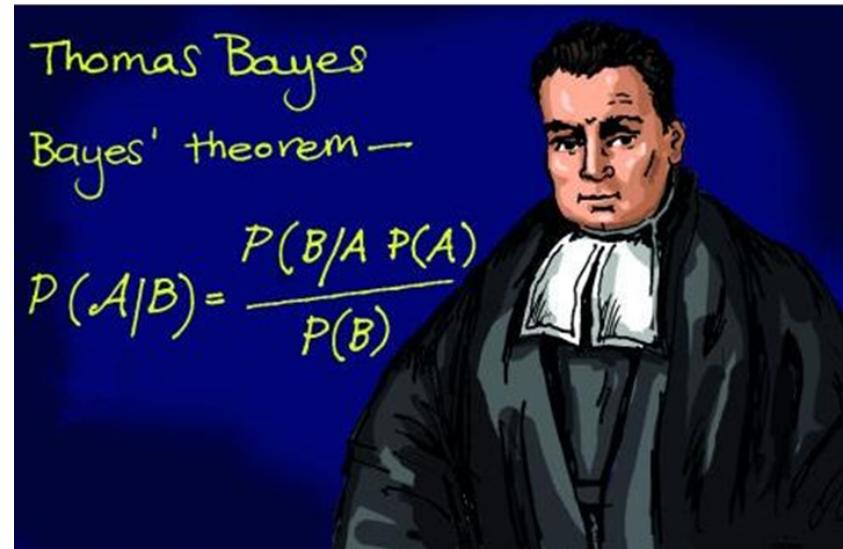
- 威廉·配第(William Petty)
- In 1690
- 政治算术(*Political Arithmetic*)一书的出版标志着统计学的诞生.

“This book does not use comparatives or superiors for thought or argument, but uses numbers to express what it wants to say, by which it examines the causes that have a visible basis in nature.” William Allotey writes.



贝叶斯(Bayes)

贝叶斯(1702-1763)

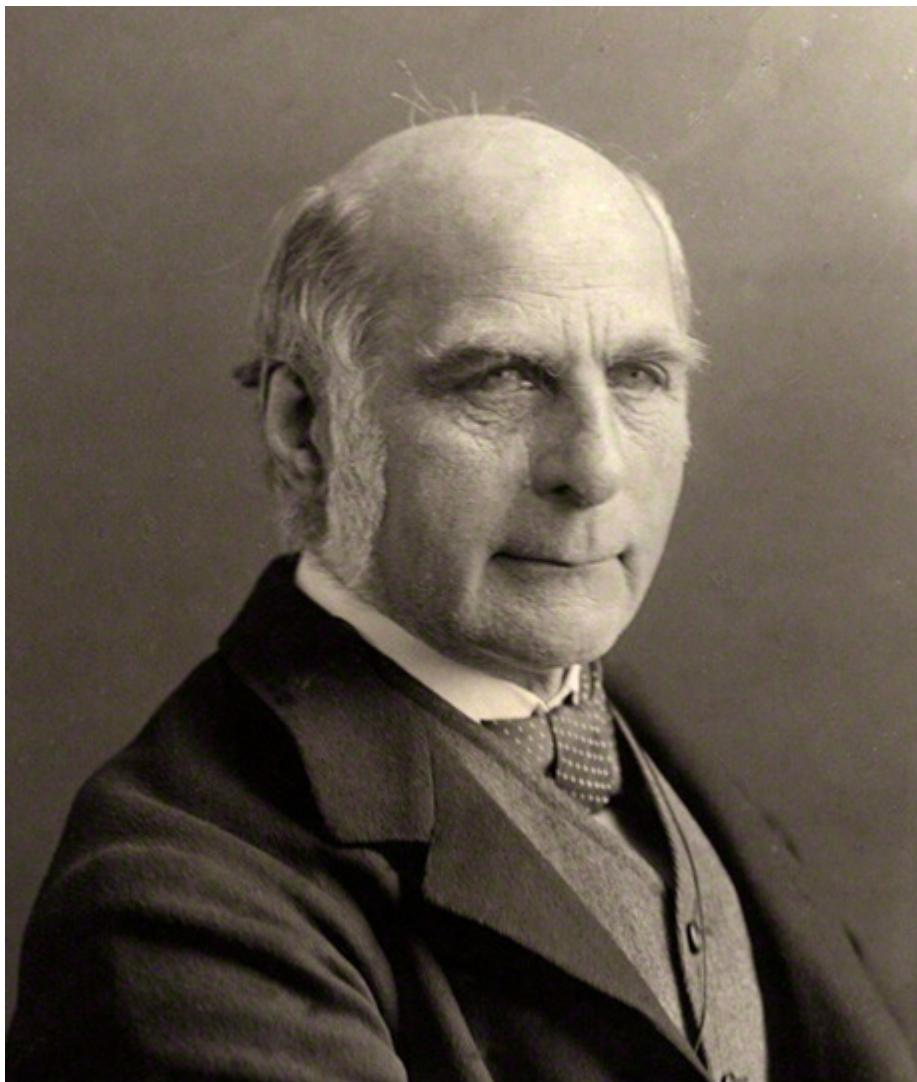


$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

1763 修建圆明园

1783 美国独立

高尔顿(Galton)



(1822-1911)

- Francis Galton
- 达尔文的表弟
- 气象学家、地理学家
- 优生学(Eugenics)
- 遗传的天才(Genetic genius)
- 《人类的才能及其发展研究》(*Research on human talent and its development*)

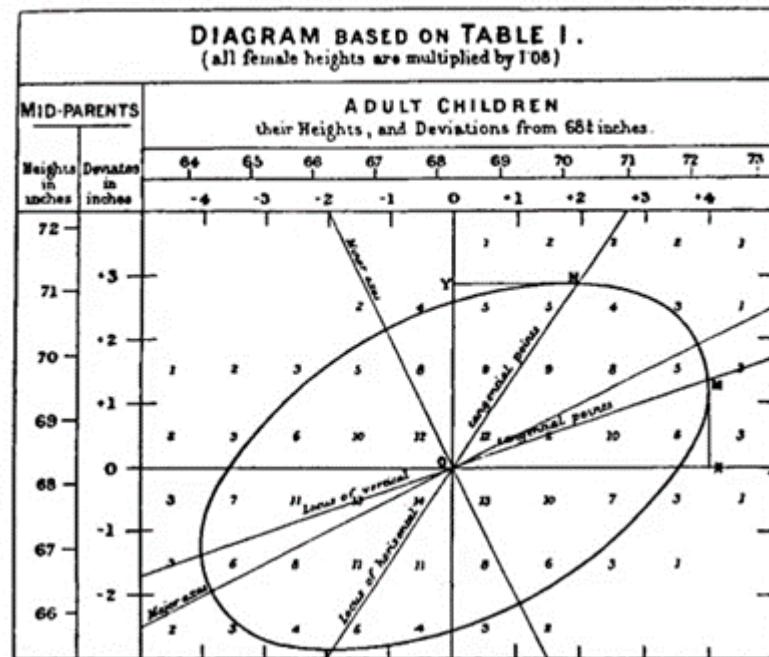
1840 第一次工业革命完成

1859 达尔文发表《物种起源》



回归分析

- 在生物统计学实验室，他收集了家庭成员的身高、体重、特定骨骼的测量值和其他特征。
- 他和他的助手们将这些数据制成表格，并对它们进行检查和重新检查。
- 通过这种方式，高尔顿发现了一个他称之为“向平均值回归”的现象。



(Salsburg, David, (2001), THE LADY TASTING TEA)



戈塞特



(1822-1911)

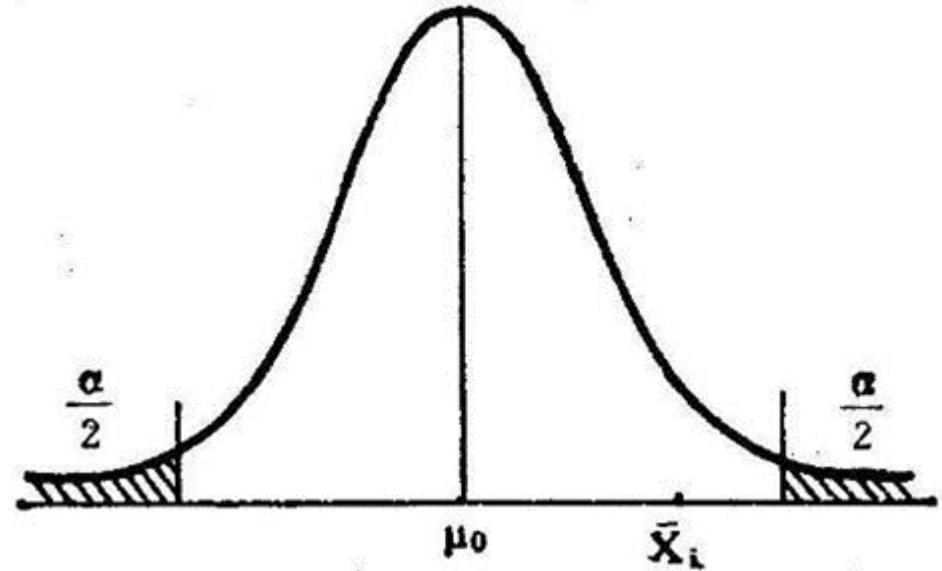
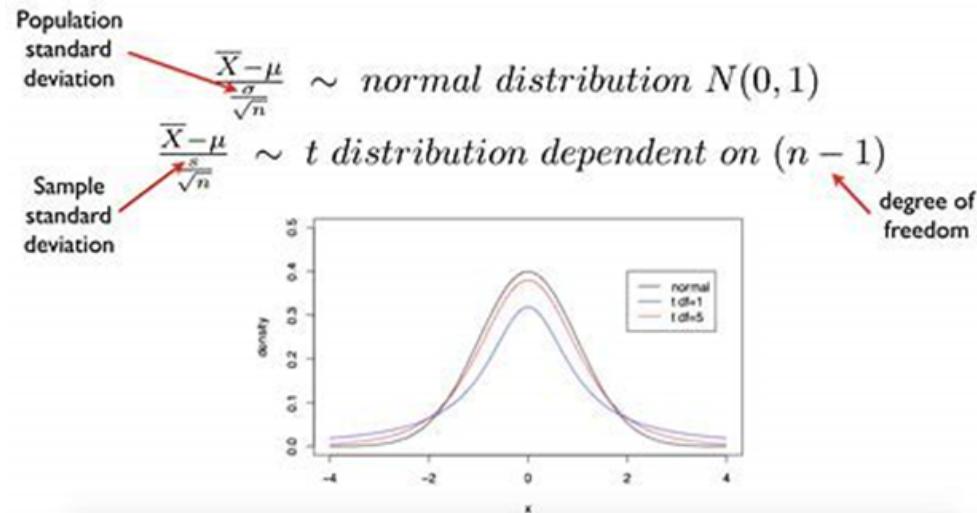
- William Sealy Gosset
- 统计学家
- 学生氏(student)
- 小样本原理

1870 第二次工业革命完成

1911 辛亥革命



检验



(Salsburg, David, (2001), THE LADY TASTING TEA)



卡尔·皮尔逊(Karl Pearson)



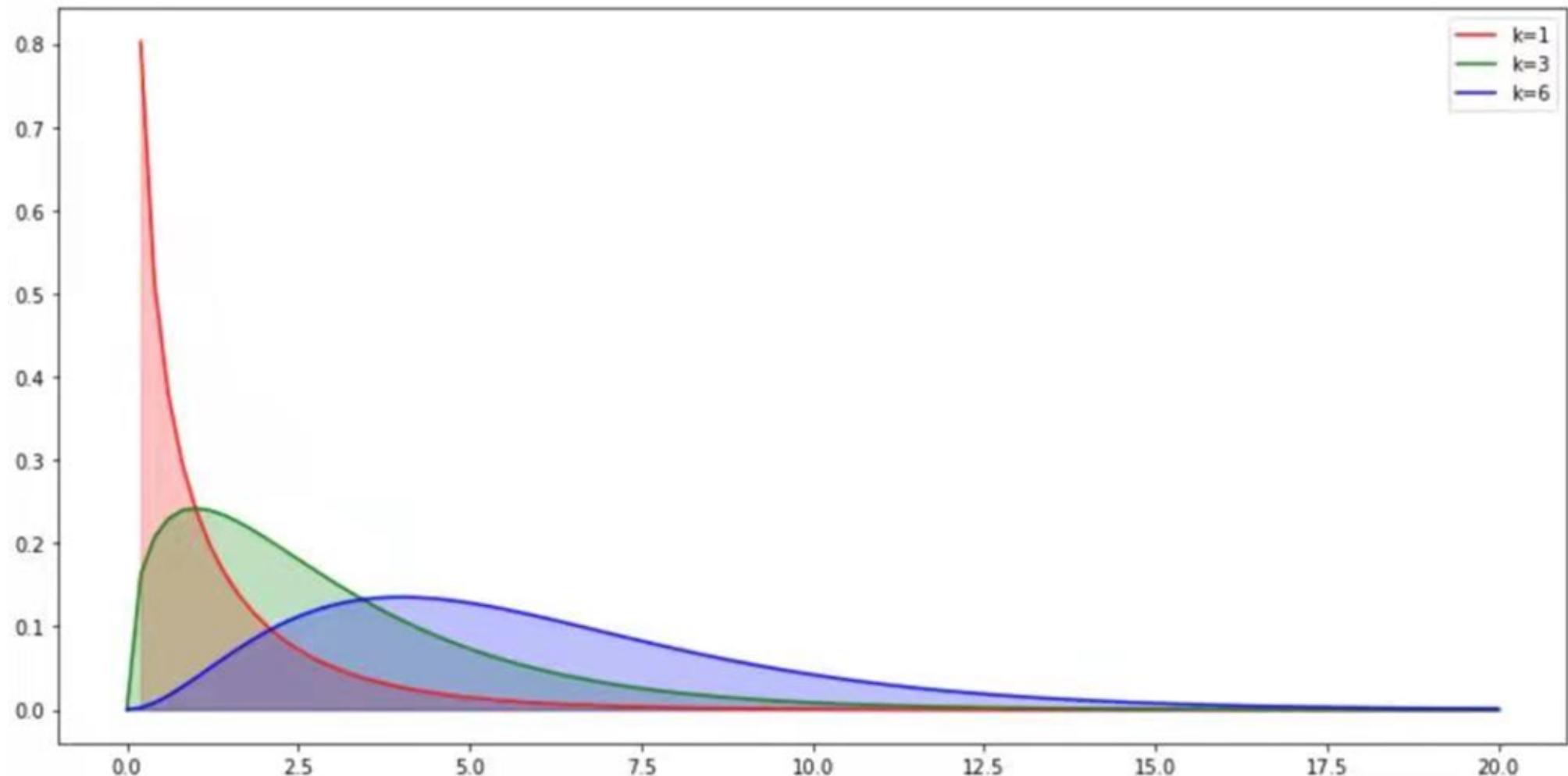
(1857-1936)

- Karl Pearson
- 数学家、生物统计学家、数理统计学家
- 现代统计科学的奠基者



拟合优度检验(Goodness of fit)

通过比较观察值(observed value)和预测值(predicted value)，皮尔逊能够使用一个统计数据来计算拟合度。



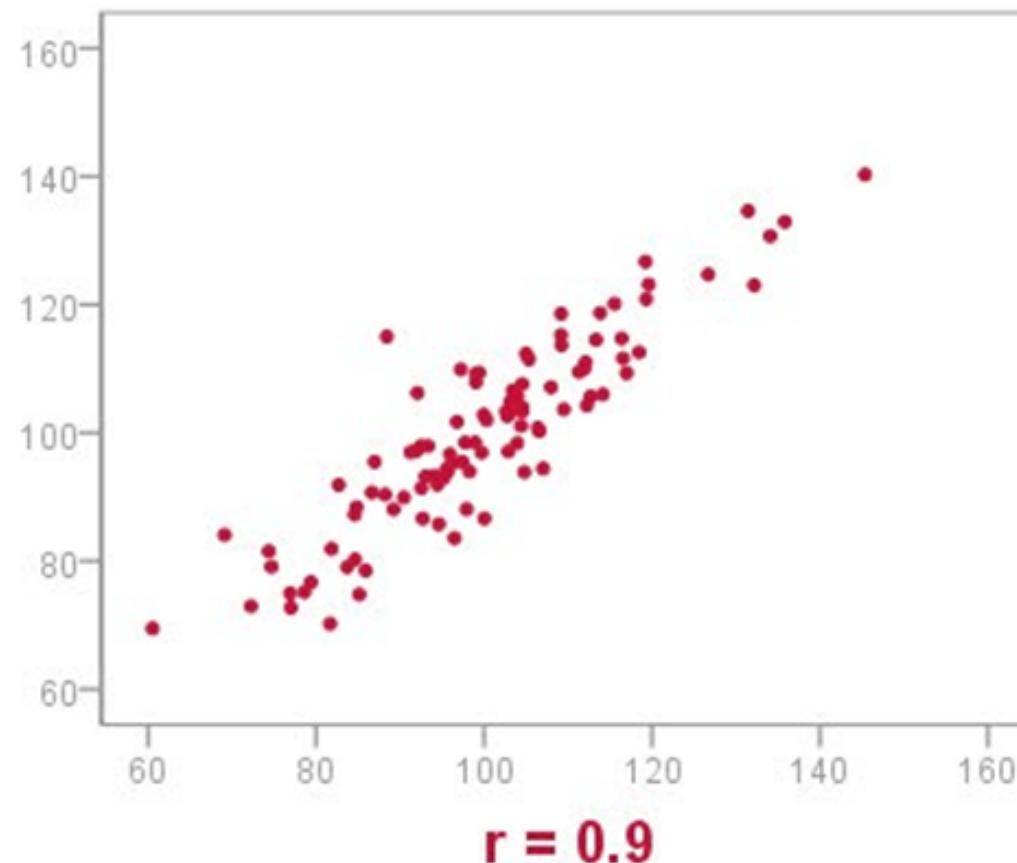
(Salsburg, David, (2001), THE LADY TASTING TEA)



皮尔逊相关(Pearson correlation)

皮尔逊相关系数用于衡量两个变量X和Y之间的相关关系（线性相关），其数值在-1和1之间。

$$p_{X,Y} = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}$$



费歇尔(Fisher)



(1890-1962)

- Sir Ronald Aylmer Fisher
- 统计学家、基因学家
- 曾工作于Lausanne农业实验站
- 《实验设计》(*Experimental design*)

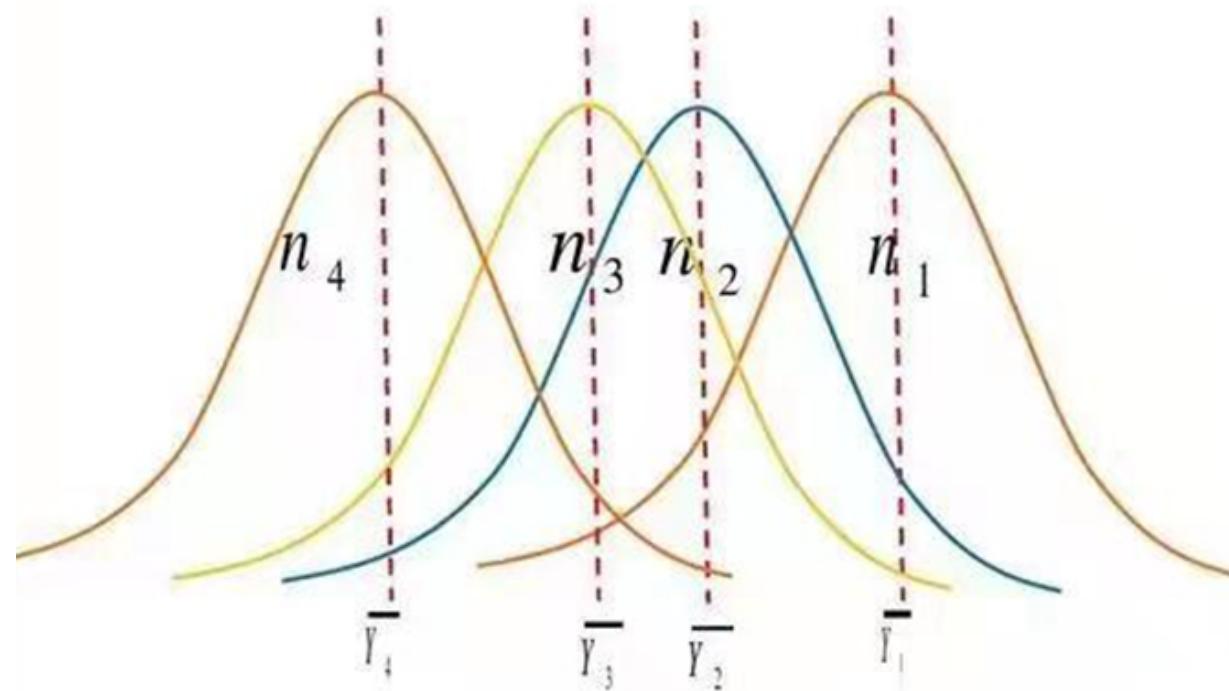
1949 新中国成立

1958 大跃进



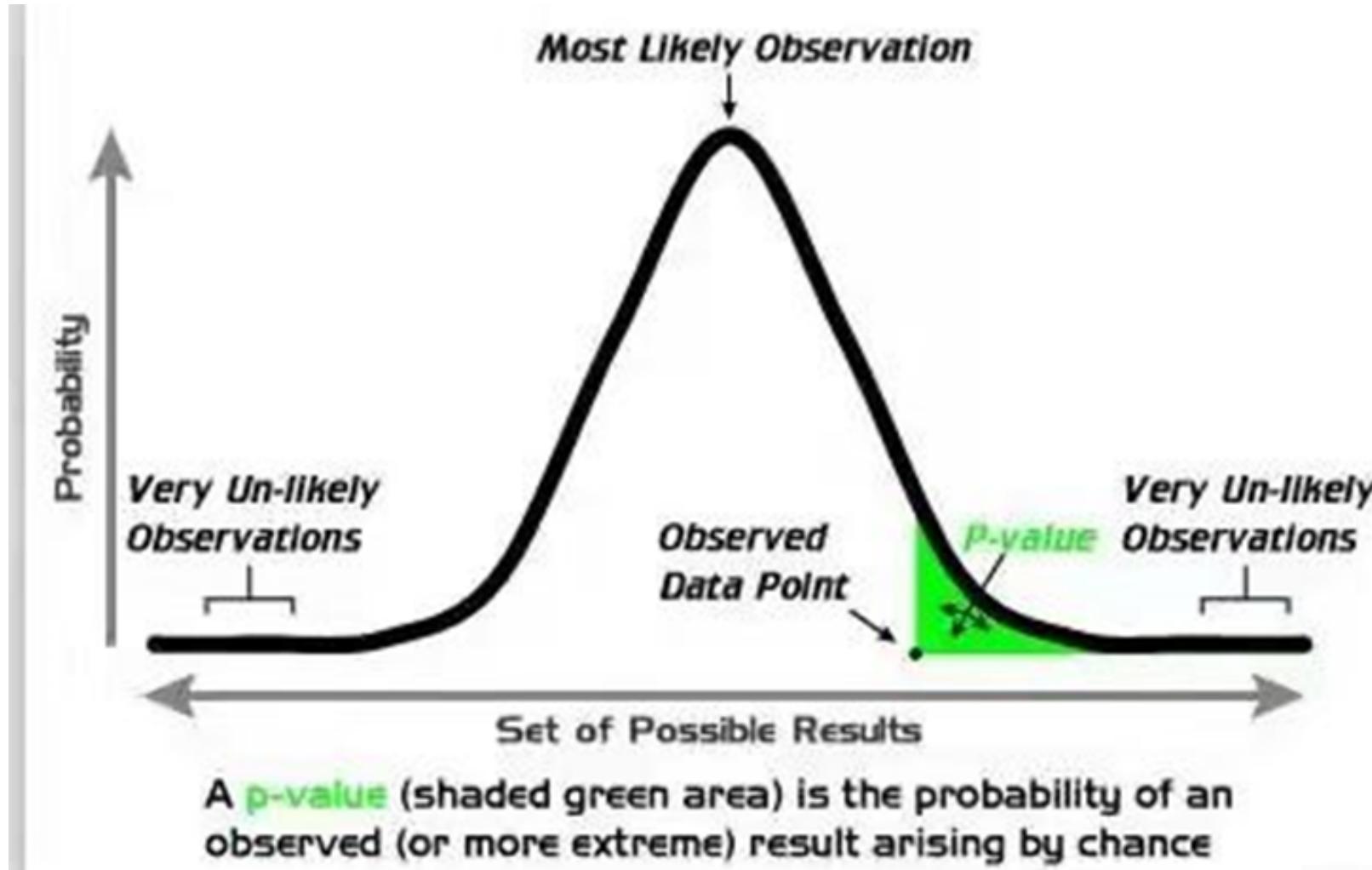
方差分析(Analysis of Variance, ANOVA)

方差分析用于检验两个或多个样本的平均值之间的差异的显著性。



p值(p-value)

R. A. Fisher开发了现在普遍使用的大多数显著性检验方法。他把允许人们报告显著性的概率称为“P值”。



(Salsburg, David, (2001), THE LADY TASTING TEA)



极大似然法(Maximum likelihood estimator, MLE)

在研究数学的过程中，Fisher意识到，Karl Pearson用来计算其分布参数的方法所产生的统计数据不一定是一致的，而且往往是有偏见的，还有更有效的统计数据可以利用。为了产生一致和有效的（但不一定是无偏的）统计数据，Fisher提出了极大似然法。

$$L(\lambda) = \prod_{k=1}^{10} \lambda e^{-\lambda x_k}, \quad \ln L(\lambda) = \sum_{k=1}^{10} (\ln \lambda - \lambda x_k)$$

$$\frac{d \ln L(\lambda)}{d \lambda} = \sum_{k=1}^{10} \left(\frac{1}{\lambda} - x_k \right) = \frac{10}{\lambda} - \sum_{k=1}^{10} x_k = 0$$

$$\Rightarrow \lambda = \frac{1}{\frac{1}{10} \sum_{k=1}^{10} x_k} = \frac{10}{9971}, \quad \therefore \hat{\lambda} = \frac{10}{9971}$$

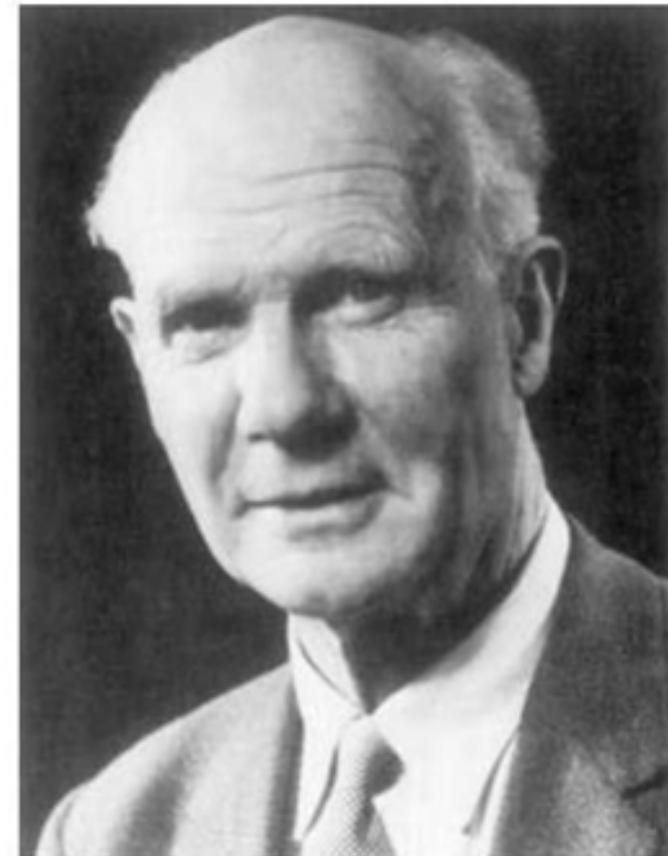
$$\begin{aligned} P(X > 1300) &= \int_{1300}^{+\infty} \hat{\lambda} e^{-\hat{\lambda} x} dx = e^{-1300 \hat{\lambda}} \\ &= e^{-\frac{1000}{9971}} \approx 0.2715 \end{aligned}$$

(Salsburg, David, (2001), THE LADY TASTING TEA)



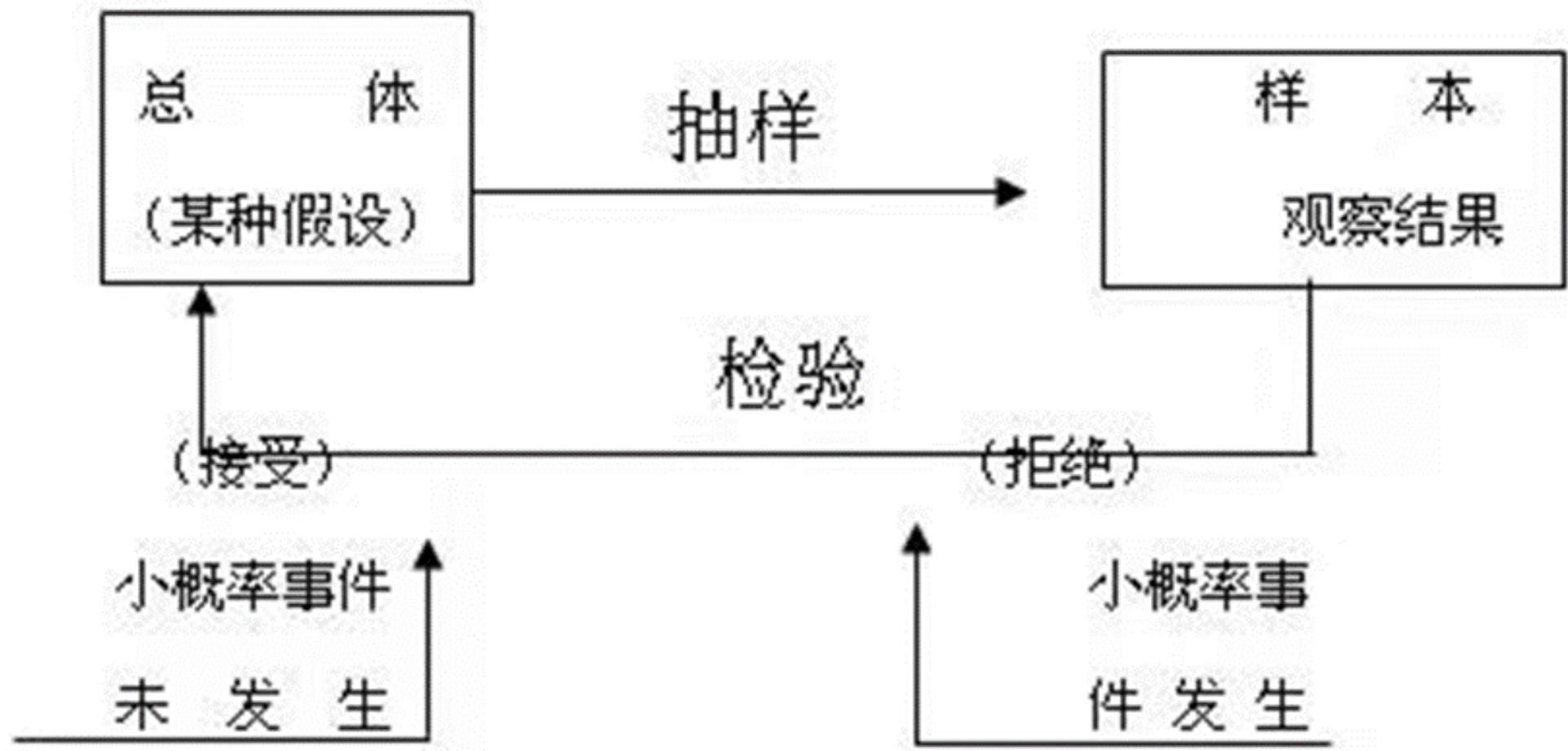
内曼和皮尔逊(Neyman & Egon Pearson)

- Neyman(1894-1981) & Egon Pearson(1895-1980)
- 假设检验
- 统计检验力



假设检验(Hypothesis testing)

为了区分用于计算Fisher P值的假设和其他可能的假设，内曼和皮尔逊将被用来验证的假设称为“虚无假设”，将其他假设称为“备择假设”。



(Salsburg, David, (2001), THE LADY TASTING TEA)



统计学简史(中国)



大案牍术



(图片来源：知乎@溜达兔)

大数据



鱼鳞图册

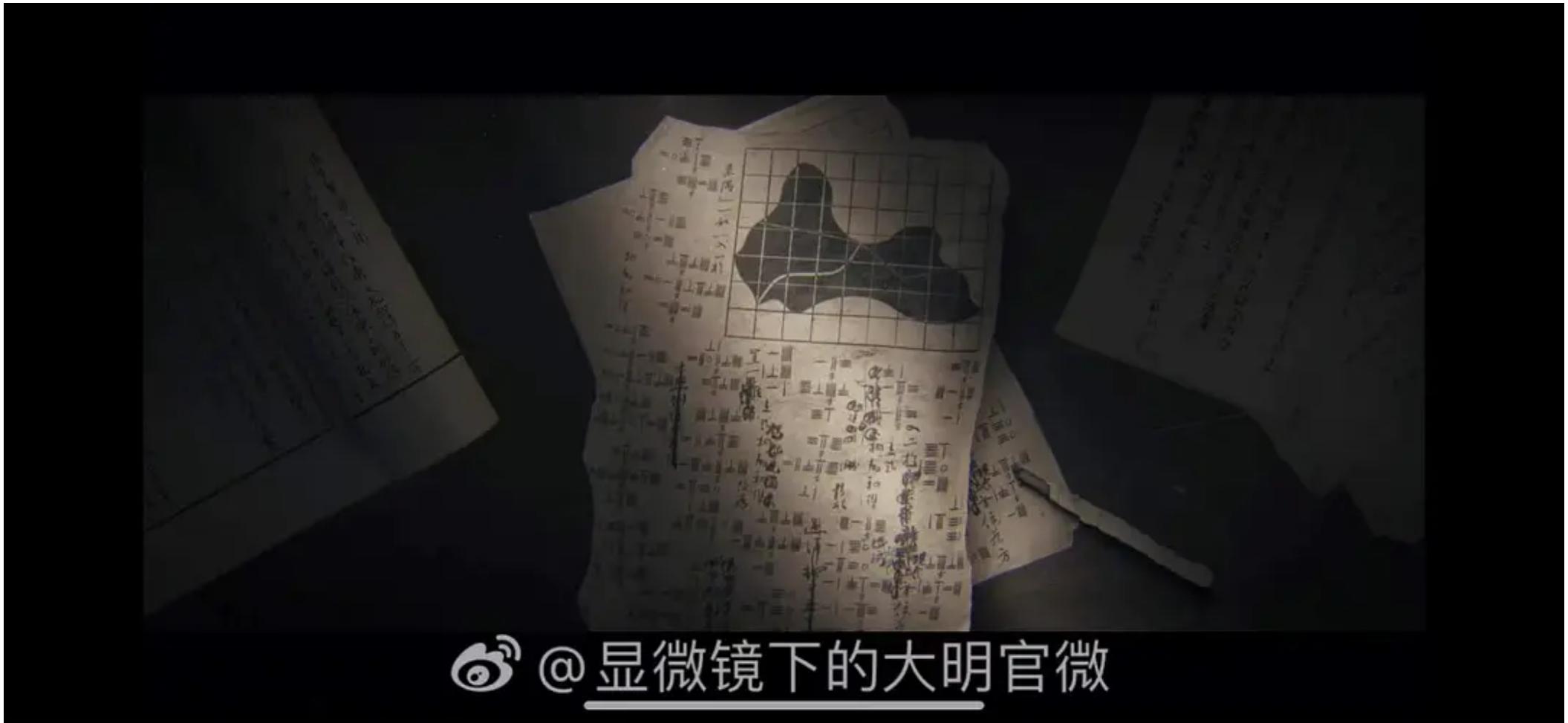


(图片来源：<https://baike.so.com/doc/6226918-6440241.html>)

大数据



推步聚顶



微博 @显微镜下的大明官微

(图片来源：微博@显微镜下的大明官微)

网格近似



上古时期

- 结绳记事
- 书契记数

夏商周

- “方以类聚，物以群分”——《周易·系辞上》
- “关石和钧”——《尚书·夏书》

春秋战国

- “举事必成，不知计数不可”——《管子·问》
- “五事”，“七计”——《孙子》
- “强国知十三数”——《商君书》



汉代

- 司马迁提出统计表编制理论
- 徐干强调人口统计是治国之本
- 仲长统重视人口调查与财经统计分析

南北朝

- 苏绰创制“计账、户籍之法”，制定“文案程式”，开创“朱出墨入”。

唐代

- 刘晏在各道设置巡院官，安排专人报告商情，采取措施，稳定物价。



明代

- 张居正对当时财政收支状况进行统计分析。他以万历年间的财政收支为依据，指出其存在的问题，并根据”量入为出”原则，提出改进措施。

清代（鸦片战争前）

- 顾炎武肯定了户籍和地籍制度的重要性。他说：“人民之丁产事业，官府必有册；土田之鳞次栉比，乡里必有图。按图以稽荒熟，为某人见业则不可隐；按册以稽某家某户占田若干、坐落某处，则税不可逋。……可见图之与册相须而不可无也。”



近代

- 吴定良(1894-1969)，生物统计学家、体质人类学家。南京高等师范学堂(后为国立东南大学)毕业，1924年赴哥伦比亚大学攻读心理统计硕士学位，1926年转到伦敦大学学院应用统计学系，师从**卡尔·皮尔逊**，1928年获得统计学博士学位。对统计学相关分析的研究和广泛应用起到了重要推动作用。
- 唐培经(1903-1988)，1927年毕业于东南大学数学系，1929年至1934年任教于清华大学数学系。1934年到伦敦大学学院，师从**内曼教授**，1937年获统计学博士学位。学有所成的唐培经，在非中心F分布对方差分析检验功效函数方面作出了贡献
- 许宝騤(1910-1970)，在国内先后就读于燕京大学化学系、清华大学数学系，1936年赴英留学，在伦敦大学学院高尔顿实验室和应用统计学系攻读数理统计，同时还在剑桥大学学习，1938年获伦敦大学统计学博士学位，1940年获剑桥大学科学博士学位。许宝騤深受**内曼**青睐，被视为内曼最优秀的学生。1938年许宝騤在**内曼**和**皮尔逊**的《统计研究报告》第二卷上发表论文《Student分布理论用于两样本问题》，在概率论和数理统计领域暂露头角，其研究方法后来被称为“许方法”，其后更以在大数定律、中心极限定理、统计推断、方差分析等多领域富有建树的突破性成果，当之无愧地成为我国最早达到世界先进水平的数学家和统计学家。



- 朱君毅(1892-1963)，浙江省江山县人，1910年考取清华学堂留美预备生。1916年秋赴美，获霍普金斯大学教育系学士学位后进入哥伦比亚大学研究所，师从爱德华·L·桑代克(Edward Lee Thorndike)专攻教育心理学与教育统计学，1922年获博士学位。朱君毅设定了一系列分析指标，利用大量的统计分析量表，对中国留美学生的学术能力、领导能力、英文能力、中文能力等四种能力经过“同伴评价”(the judgment of associates)，得出“成功”的结果，将该结果与高中成绩、大学成绩、个人情况、留学时间——加以对比，得出了四种能力之间的相关性以及与留学时间之间的相关性，进而对国内中英文能力的培养、留美期限、留美生出国前应具备的学历学识等提出了建议和主张。**曾就职于南京女子师范学校**



- 张厚粲(1927-2022)，1927年生于北京。1948年，张厚粲先生毕业于北平辅仁大学心理学系并且留校任教，1952年经过高等院校教育的调整转入北京师范大学，现任北京师范大学心理学院教授、博士生导师，校委员，中国心理学会常务理事，曾任国际心联副主席。2005年台湾辅仁大学授予张先生心理学博士学位。张厚粲先生的主要研究领域包括实验心理学、人类的认知以及心理和教育测量。她首先将认知心理学引入中国，在心理统计和测量方面，她主编了我国最早的心理与教育统计学的教材，率先开设了心理测量的课程，组织修订中文标准推理测验、韦氏智力测验，编制中国儿童发展量表，使得测验在我国广泛地应用和发展。



- 古希腊时期（约）
 - 亚里士多德 统计学起源

- 1690
 - Political Arithmetic 统计学正式诞生（国势学派&政治算术学派）

- 1890以后
 - 现代统计学科创立（卡尔·皮尔逊, 费歇尔, 奈曼, E·皮尔逊等）

- 汉代以前
 - 统计学相关思想萌芽
- 汉-清（鸦片战争前）
 - 统计学相关思想在国家管理上得到应用

- 1690
 - 中国处于康熙皇帝时期

- 近代（鸦片战争后）
 - 现代统计学思想引入

- 1895
 - 甲午战争失败
 - 一批中国学生前往海外学习统计学



统计学的发展



chatGPT



(图片来源：<https://americanfaith.com/users-figure-out-how-to-break-chatgpt-to-ignore-ais-leftist-bias/>)

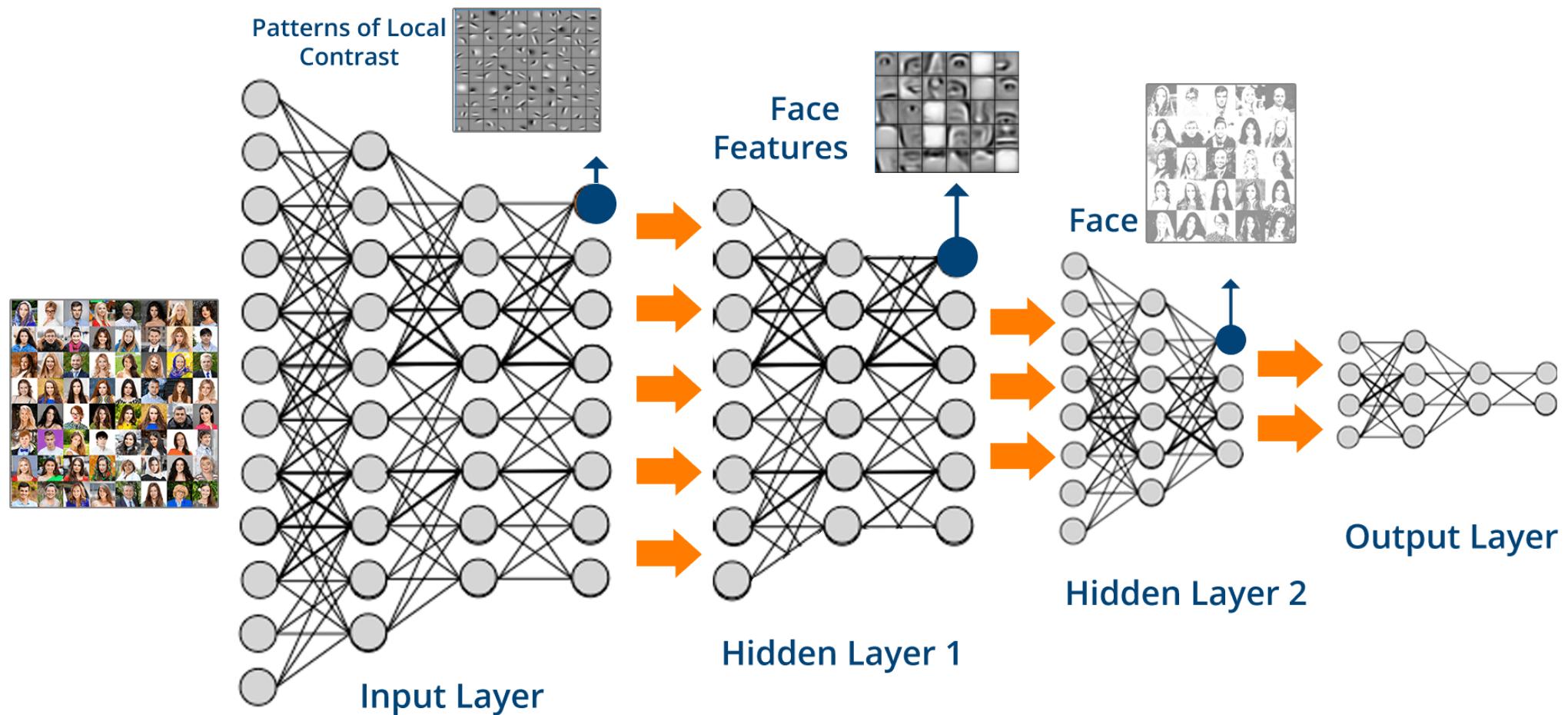


AlphaGo



(图片来源：<https://medium.com/@greychen/alphago-research-review-6e8632221a22>)

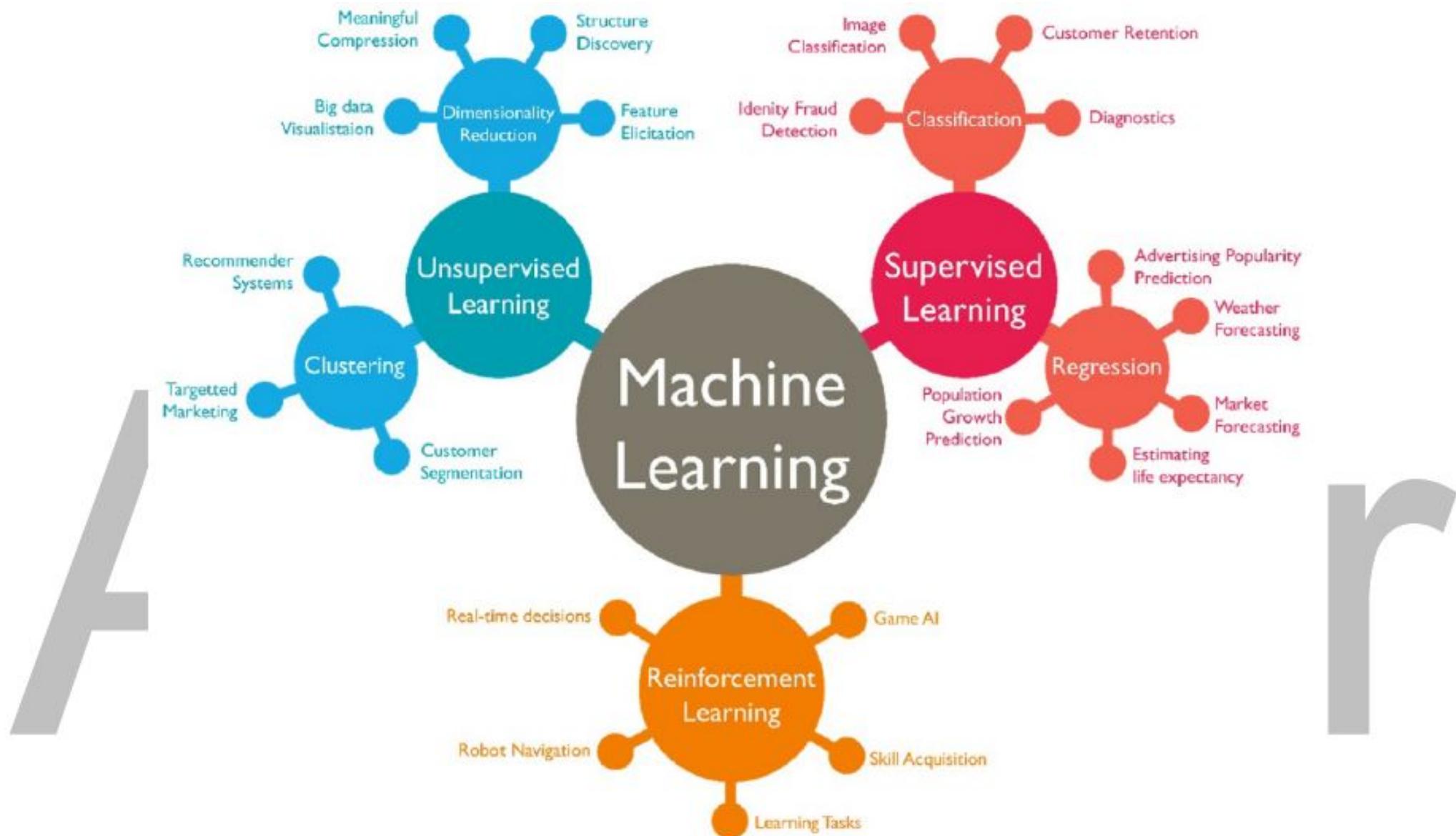
Deep learning



(图片来源: <https://cn.bing.com/newtabredir?url=https%3A%2F%2Fwww.skampakis.com%2Fwhat-deep-learning-is-and-isnt%2F>)



Machine learning



(图片来源: <https://auto.ifeng.com/qichezixun/20200120/1374691.shtml>)



生活中的统计学

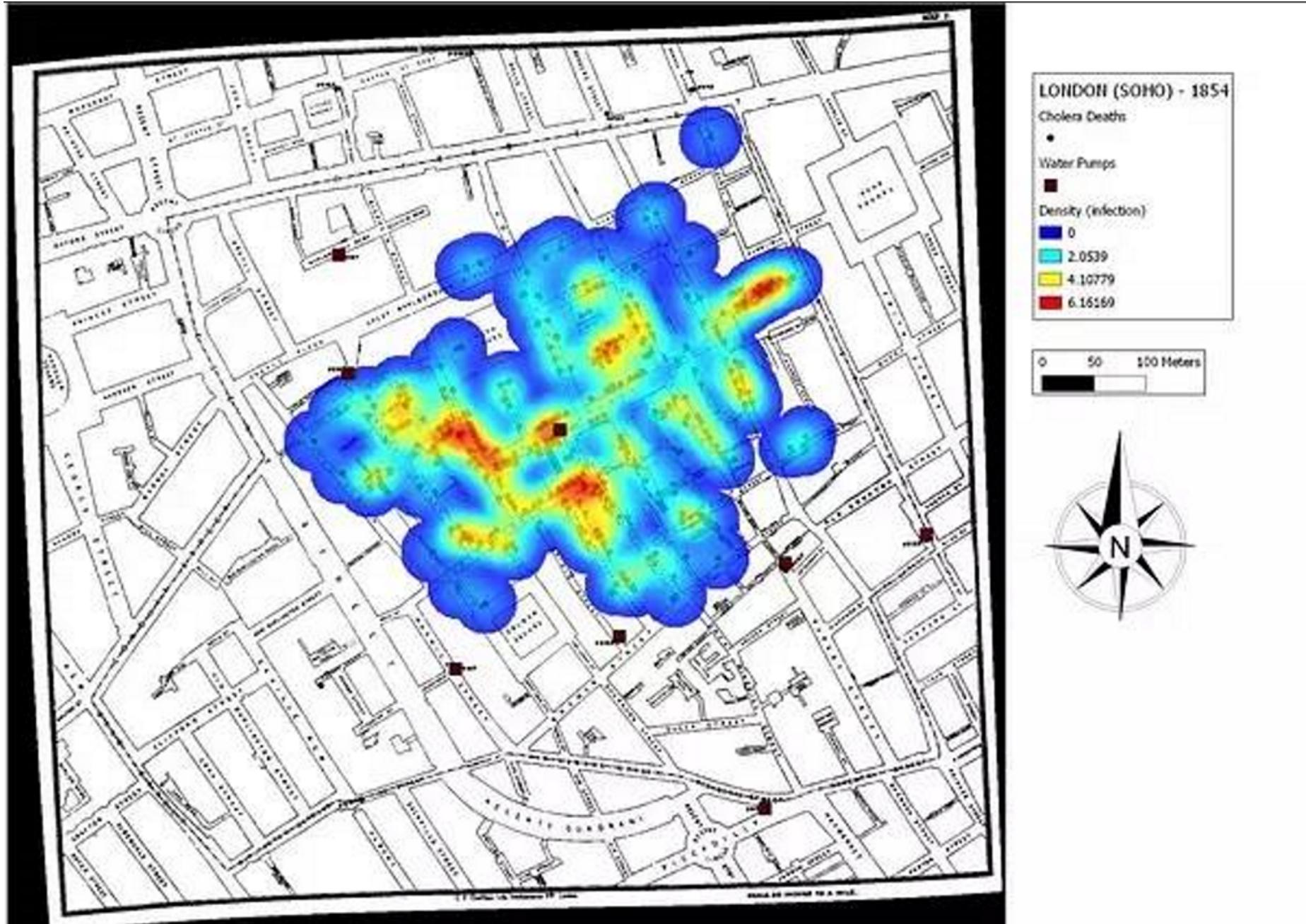


统计与霍乱(Cholera)



- 霍乱(Cholera)是一种急性腹泻疾病，由肠道感染霍乱弧菌 (Vibrio Cholerae*) 引起。O1或O139。
- 据估计，全世界每年有290万个病例和9.5万人死亡。这种感染通常是温和的或没有症状，但也可能是严重的。
- 霍乱最容易在水处理不当、卫生条件差和不卫生的地方发生和传播。

(资料来源：Centers for Disease Control and Prevention, USA)





Broad street is still in London



统计与新冠

CDC: Face Masks Don't Prevent COVID-19, Study Finds
Masks Have Negligible Impact On Coronavirus Numbers

Stanford Study Results: Face Masks are Ineffective to Block Transmission of COVID-19 and Actually Can Cause Health Deterioration and Premature Death

NEWS

Ron DeSantis Says 'Not Healthy' for Students to Wear Masks Amid New CDC Guidance

BY JULIA MARNIN ON 7/29/21 AT 9:50 AM EDT



0:03 / 12:36

Why Face Masks DON'T Work, According To SCIENCE

published in Annals of Internal Medicine.

In the largest randomized controlled trial to date w/ 6,024 subjects, medical masks were ineffective protection against

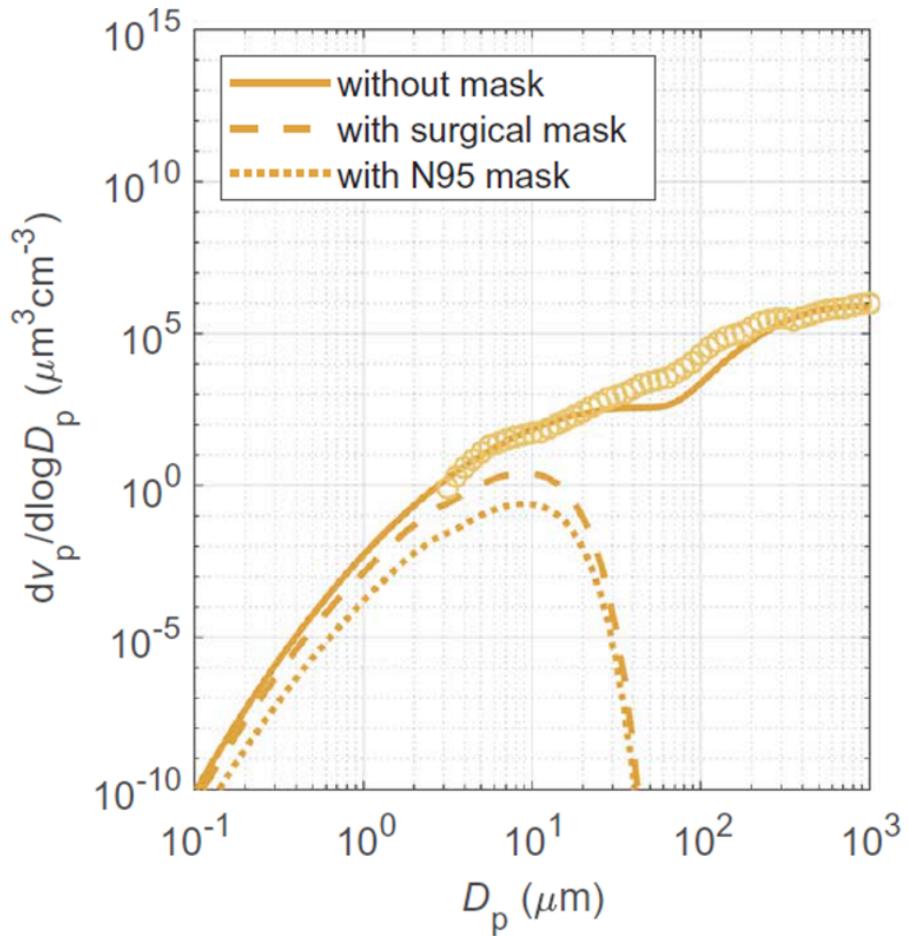
acip.journals.org/doi/10.7326/M2...

10:32 AM · 18 Nov 20 · Twitter Web App





What evidence supports the use of face masks
against the coronavirus?



v_p : 体积浓度

D_p : 呼吸道颗粒的直径

$dV_p/\log D_p$: 体积分布函数

(Cheng, Y., (2021), *Science*)



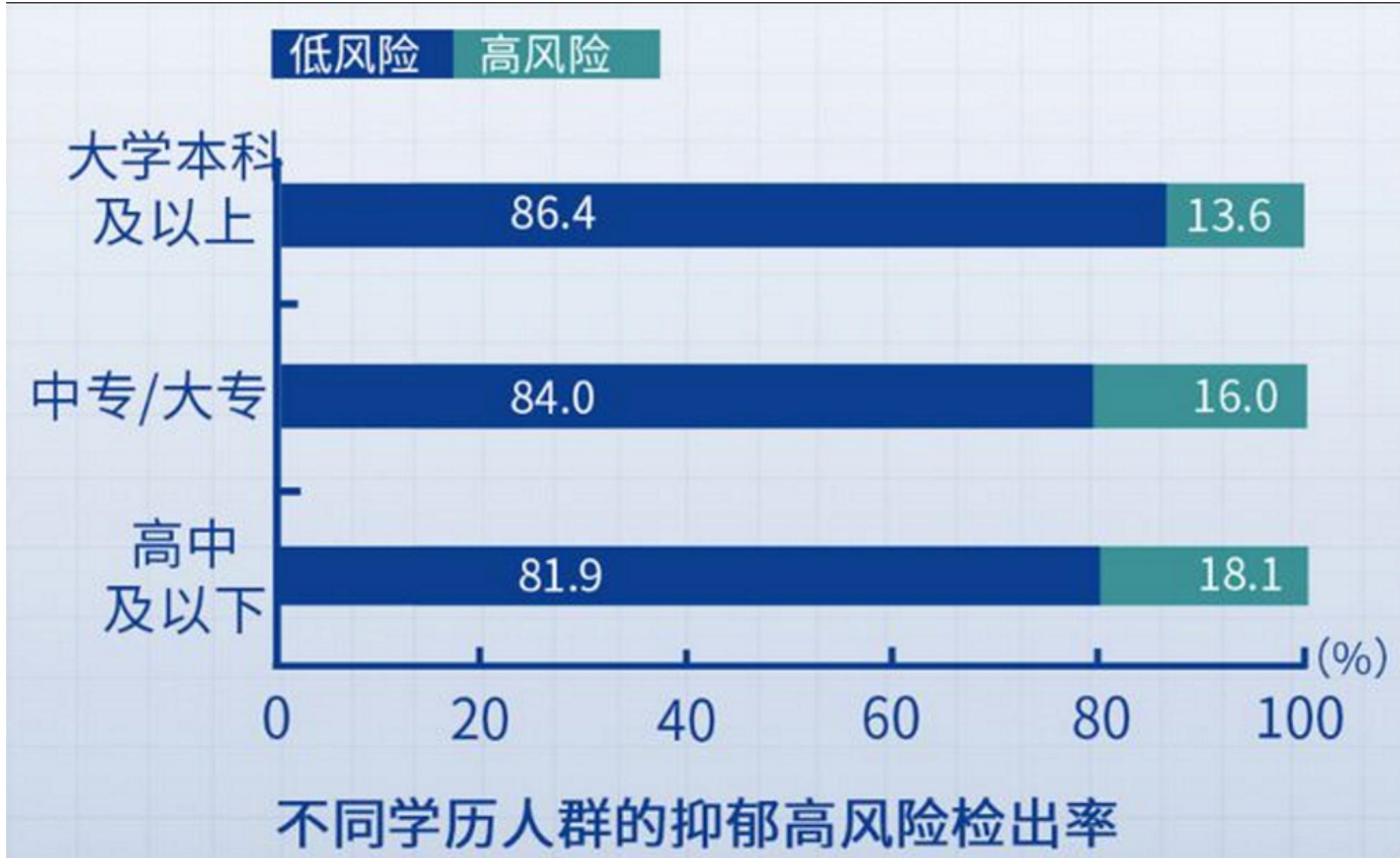
心理健康

	无抑郁	轻度抑郁	重度抑郁
2020 年青少年样本	75.4	17.2	7.4
2009 年青少年样本	75.8	16.8	7.4

(资料来源：《中国国民心理健康发展报告（2019～2020）》)



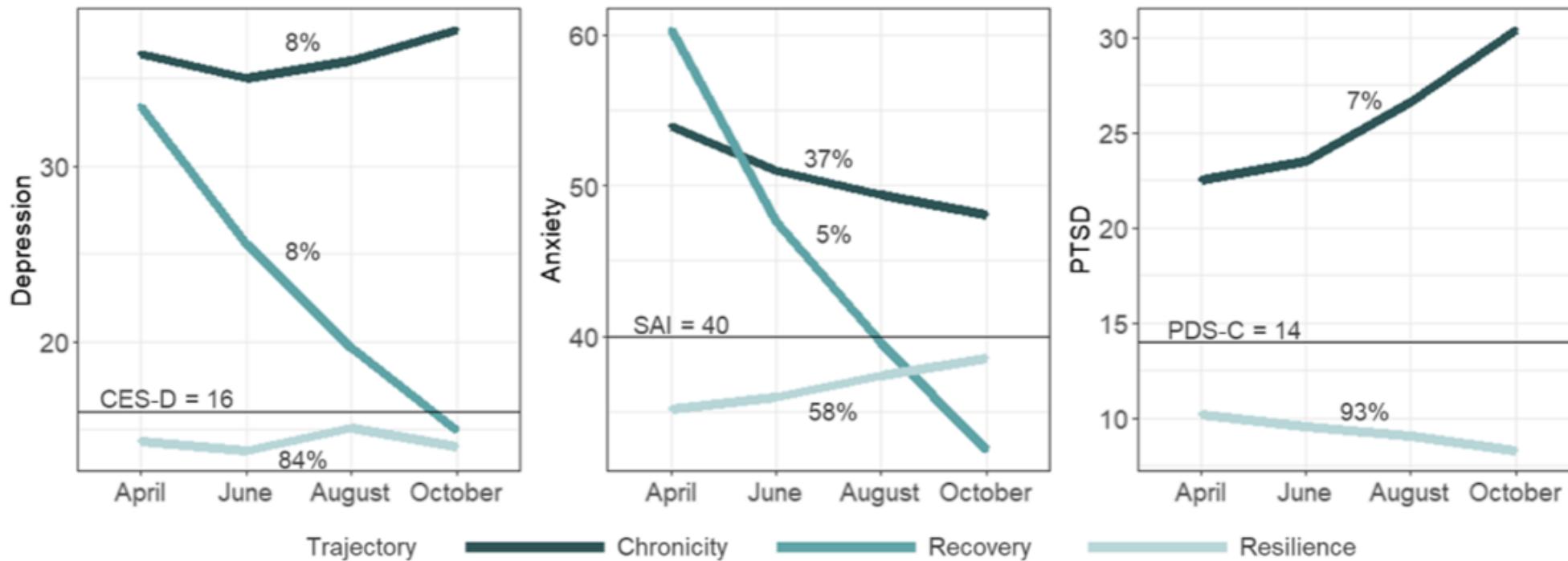




(资料来源：《中国国民心理健康发展报告（2019～2020）》)



武汉人民心理健康的纵向趋势



(Shuquan Chen, (2022), *American Psychologist*)

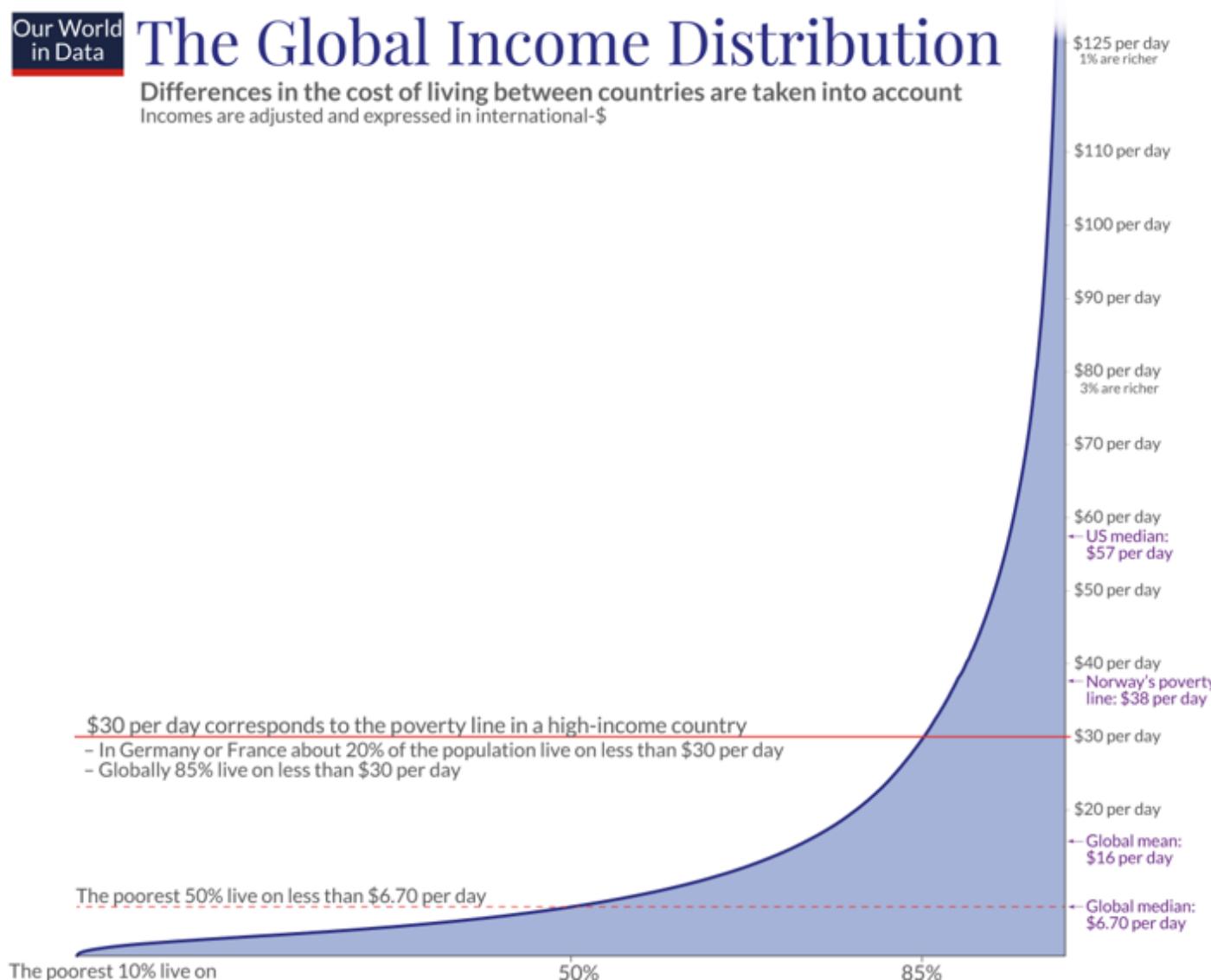


统计与收入不平等

Our World
in Data

The Global Income Distribution

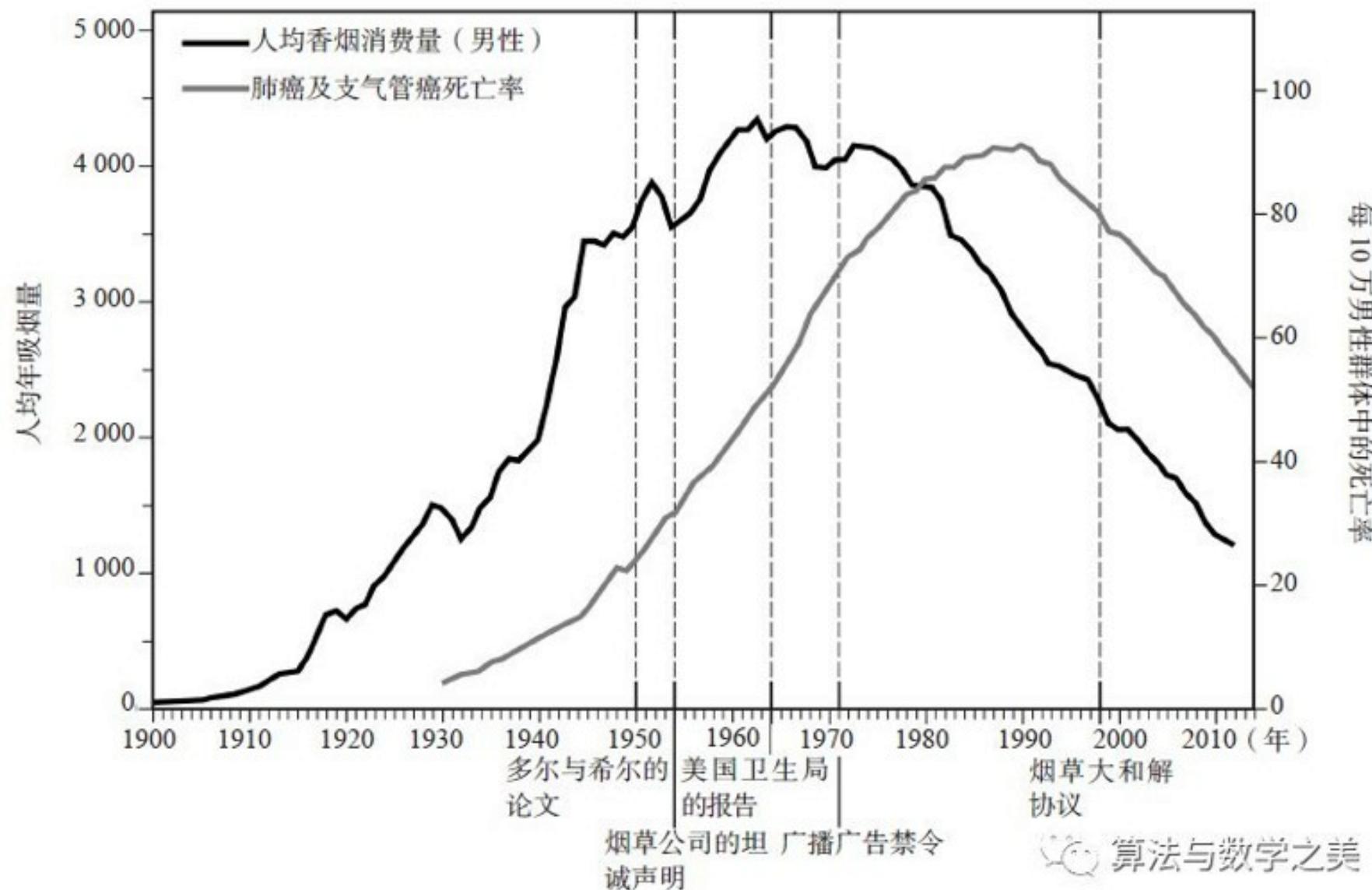
Differences in the cost of living between countries are taken into account
Incomes are adjusted and expressed in international-\$

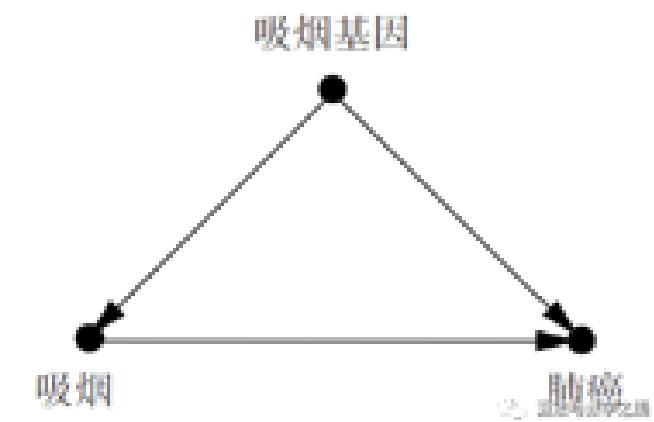
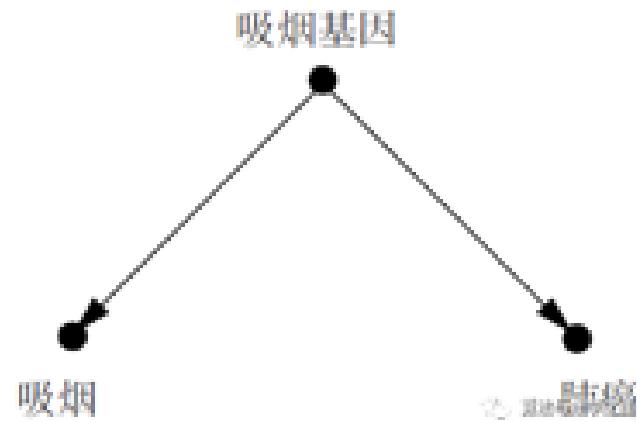


(Max Roser, (2021))



吸烟会导致肺癌吗





(资料来源： Judea Pearl & Dana Mackenzie, (2019), 为什么：关于因果关系的新科学)



总结

课程概况

统计简史

- 统计学的起源
- 西方统计学史
- 中国统计学史

统计在生活中的应用



