

现代科研指北

于淼

2018-07-11

目录

1	前言	5
2	科研在搞什么鬼	7
2.1	科研老鸭汤-科学哲学沿革	7
2.2	科研职业化-问题为导向	10
2.3	科研精细化-体面的博士	11
2.4	科学知识的五个层次	12
2.5	知识体系的时间构建	13
3	科研现状概览	15
3.1	国内版	15
3.2	国际版	15
3.3	趋势	15
4	思维工具篇	19
4.1	科学思维	19
4.2	模型思维	19
4.3	统计思维	20
4.4	估算法	20
5	实验	21
5.1	实验设计原则	21
5.2	定性实验	21
5.3	定量实验	21
5.4	思想实验	21

6	数据处理	23
6.1	多重比较	23
6.2	多重检验	23
6.3	回归	23
6.4	预测	23
6.5	仿真	23
6.6	可视化	23
7	文献	25
7.1	文献管理	25
7.2	信息收集	27
7.3	文本挖掘	27
7.4	荟萃分析	27
8	学术生活	29
8.1	项目管理	29
8.2	学术出版	29
8.3	学术会议	31
8.4	审稿	31
8.5	学术合作	32
8.6	讲课	32
8.7	课题组管理	32
8.8	学术声誉	34
8.9	学术道德 / 伦理	34
8.10	案例	34
	附录：现代科研兵刃谱	35
	文本编辑	35
	数据处理与绘图	35
	学术交流	36
	审稿	37

目录	5
数据分享	37
代码管理	37

Chapter 1

前言

才疏学浅，不知何为真，仅通少错之法，故不敢言南，仅指北。或曰：现代科研挖坑 / 跳坑指南

Chapter 2

科研在搞什么鬼

2.1 科研老鸭汤-科学哲学沿革

做科研一般都不讨论哲学，太多形而上的东西，说也说不清道也道不明还无法证伪。但懂一点科学哲学还是很有必要的，不然很容易研究着研究着就会觉得自己做的东西是垃圾，是谋生的工具，虽然从某个角度看也没错，但科学哲学无疑是应对这种心态最好的老鸭汤。

2.1.1 古希腊

哲学是爱智慧这个梗就不多说了，扯古希腊也显得俗套，反正有了古希腊人才有了理性跟逻辑的提法。古希腊前面的历史可理解成经验性知识的发展，知识多了就要有规律总结出来，逻辑和理性可看作用来生成规律的知识。其实哲学就是认识世界的知识，泰勒斯有一套，毕达哥拉斯有一套，赫拉克里特有一套.....大家能自圆其说就来一套，对不对另说，不服就辩论，赢了就是真理，输了就是谬误。如果说逻辑与理性出自于这些街头巷尾的辩论我一点也不会奇怪，因为两套理论对比，总得有两方都认可的法则才有结果，理性或许就是这种普遍性知识的产物。

不过辩论有诡辩这一说的，苏格拉底看不下去了就说你们这些人都觉得自己对，但有可能是不对的，反正我自知我无知（这句是我认可最有智慧含量的句子，还有一句：天下没有免费的午餐）。老苏不怎么关注解释万物，有点回归个人或社会的意思。到了柏拉图直接就理想国了，世界形物均为理型的影子。再到了亚里士多德就不怎么废话了，直接取消理型世界的存在，认为万物有因，这个因就是所有问题的因，寻找到最终因，真理就明了了。这货还不知道这个看法后来发展成第一推动问题，宗教界觉得只有全能的主有这能耐，就把亚里士多德的理论吸引到宗教哲学里去了。

同时，我们现在所提到的科学源于日本，可理解为分类的知识，而最早对人类知识体系分类的就是亚里士多德。他还很神奇的将自己的目的论揉到这个分类里去了，所以这个体系很完整，能解释的东西很多，所以后来几百年大家就都用了这个体系。其实这时候科学知识更适合分到亚里士多德所谓的自然哲学这个科目里，这个科目特指自然现象的规律及探索方法。这里需要注明的是数学更多是工具，数学化不一定就

代表科学，另一个需要注意的是逻辑学，这货的三段论十分精彩，以至于要不是后来哥德尔横空出世，任谁也动不了根基。

2.1.2 中世纪

中世纪黑暗吗？如果看天气应该跟现在差不多，但这个黑暗的印象大致源于天主教对知识的垄断，而知识也反过来服务宗教，而宗教理性在一定程度上促进了我们对世界的认识，所以盲目对立宗教跟科学没必要，很多前期知识都是不少富有宗教热情的理性人士总结的。只不过知识很多种，科学在那年代连个独立的名字都没有，所以你看，牛顿写本书叫《自然哲学的数学原理》，跟现代意义上的科学没啥关系，所以你管他信仰什么呢。这个时候，科学知识跟形而上学还是分不开，很多知识有严谨的数学形式但你无法证实，很多天文学知识就这德行，你去看看托勒密体系，圆环套圆环的也能解释现象，那哥白尼的日心说为什么就流行了呢？因为是真理？因为结构简单？还是因为他用了别人看不懂的语言写出来的？或者说反对者死绝了新理论就流行了？总之，没有实证的理论的流行不会是你想的那么简单，但一般来说，人们都喜欢简单且解释面广的理论，宗教也这样，毕竟美的东西都是上帝赐予的。在科学这个提法之前，用一套知识来解释世界是各代学者所向往的，才不管验证什么的，理性重于事实。

2.1.3 1500 年以后

时间点不太好找，但历史的发展是伴随知识的增长的。大航海时代为人类的知识提供了一个海量来源，文艺复兴带来了人性的解放，宗教改革让生活走出了政教合一，总之，经验开始比逻辑更为人接受。最开始是欧洲大陆的理性说与英国的经验论的争执，争论核心在知识的构建是从理性出发还是从经验出发，这两种观点打架几百年，到了 19 世纪末大家都不争了。因为实证主义一统江湖认为从经验中提取逻辑，然后再证实就 OK 了。这时候科学哲学才独立出来，而观察式的经验也开始让位于实验式的事实，人们不满足于被动接受知识，开始主动去寻找真相。

2.1.4 逻辑实证主义

当人自己把握了主动权，原有的常识知识就要被逻辑重新检验，而无法检验的就划到形而上学这一类里留给做宗教神学的人去讨论。换言之，从柏拉图开始的将现实世界与理想世界的区分被打破了，原来的哲学家都醉心于构建理想世界而不关心现实生活，而逻辑实证主义则要求通过生活的事实来寻找真相。换句话说，经验事实及逻辑推理被结合用在真理的探索上了。而经验事实的崛起则伴随着归纳法的崛起，事实成为知识的唯一来源，科学开始渗入并改造哲学方法论，这一转变真正让科学有了真理探寻的光环，一举扫清神秘主义与宗教束缚，直到今天还在深刻的影响着每一个科研工作者。

2.1.5 否证主义

但不久大家发现不对头，因为归纳法不如演绎法严格，得到的结论有局限性，不够严谨。这时候波普就说了，演绎法靠谱！大家都提假说，然后验证它，出现反例就把假说否了，不能否证就不科学，这就是证伪。一时间大家都接受了，神马佛洛伊德，历史唯物主义都因为自洽但不能证伪给踹出科学圈了。

不久又有人感觉不对了，一方面演绎法很难产生新知识，另一方面貌似假说是无穷无尽了。证实比较费事，证伪容易但很多理论就垮了。为了调和这个矛盾，否认主义给出的答案是演绎法虽不能产生新知识，但假说的产生不是无缘无故的，而知识的进步应该通过大胆猜想的确证与谨慎猜想的否认来完成，一个推翻的理论必然联系着新理论的提出，这时不断发展的，而科学的任务就是处理进步问题而非回答真理问题。形而上学也并完全被排斥了，因为假说的提出有时就是没有事实证据的。进一步讲，波普尔将世界分成世界 1，也就是物理世界，世界 2，也就是精神世界，然后又分了个世界 3，也就是客观知识世界。这种三分法其实是将柏拉图的理念世界进化了，同时也留下了世界 2 的个人空间。每个世界都在进化，这就是科学发展的轨迹。一口吃不成胖子，我们就去试错吧！猜想与批判这一否认主义的核心思想也是当下科研中比较闪光与巧妙的实验设计动机来源。

2.1.6 历史主义

前面那些理论的提出者大都数理化出身，推理证明构建系统很在行，但没案例不成啊，得解释得了现象啊。其中一些人翻了翻了史书，发现很多发现不是通过证伪得到认可的，也不是建立在大量归纳的基础上，而是具有“历史性”。也就是逻辑不怎么灵光，然后他们就说咱以史为鉴吧！拉卡托斯就搞出了个硬核软核的理论，大意说一个理论是有生命力的，硬核部分无须质疑，有保护带，一时半会死不了。需要缝缝补补的是外围软核，什么时候硬核也不行了，就退出历史舞台了。这个解释保全了科学理论体系，也就是堵了民科的路，要知道民科最喜欢证伪，一个错误就否了整体，现在拉卡托斯说不成，得慢慢来，有历史的。

不久又有人感觉不对了，我怎么知道现在的硬核到底对不对？拉卡托斯这时就呵呵了，交给历史评价吧！库恩在这个背景下提出了范式，他本身有较强的历史功底，手头案例多，所以有了科学共同体这个说法。大意就是一个时代的真理主流说了算，这伙人挂了而接任的更多采取了另一种解释现象更多的理论，那这个理论就上位了，就革命完成了。前面那个时期比较压抑就叫前科学，后面上位了就是常规科学。之后又有新现象解释不了了就有了危机，这时候新理论又出现了，再搞一次革命就 OK 了。范式是用来区别前科学与常规科学的，范式通常是一套当前时代科学共同体所使用的理论体系，而这个理论体系要比之前的更能解释更多的问题也更严格。这理论比拉卡托斯那一套通俗易懂，那年代搞政治的一看有革命二字纷纷表示深有体会，大力推广之，所以范式着实火了好一段时间。

库恩的范式革命是格式塔式的转换，历史上一共也没发生几次，真正有益的是他对范式定义时要求要有自称科学的学科要有自己的理论体系与假设且对现实世界产生作用，这个理论自身并不要求科学家的态度是客观的，但范式自身要是客观的。这时候，大家都不愿搭理真理性这茬了，因为都清楚对错问题是历史性的。同时范式也把形而上学彻底请回到科学体系中并认为对科学的发展是有益的，要知道波普尔虽然不拒斥形而上学但本质还是批判形而上学的。所以历史主义的强调使得真理相对化。

2.1.7 无政府主义

事实上你沿着这个思路走下去发现貌似科学发展跟三国演义差不多，不在于对不对而在于认可的多不多，有没有跟你闹革命的。当然因为实证主义的余威，理性与逻辑在科学研究中是绕不开的。这时候来了个更霸气的费耶阿本德，一拍桌子，科学跟别的知识没啥区别，不能特殊对待。后来流传到世上的就是那句 anything goes，很多人认为这货终结了科学哲学的发展。从 20 世纪初到六七十年代这个学科就完蛋了，这就是科学哲学的学科危机。

2.1.8 实用主义

逻辑委实打不过历史，原来那些搞科学哲学研究的还没死就没饭碗了，生存是硬道理。他们发挥了科学共同体的作用，把费耶阿本德斥为异类、后现代。但他说的话又绕不过去，这时候蒯因跳出来科学哲学还要发展，不能 anything goes，科学不科学总得有个标准。美国人想来想去想到了有用两个字，然后大家纷纷鼓掌。理性，历史都打不过生存这个命题。有用是硬道理，有用解释一切，然后就没有然后了。

2.1.9 其他

除此之外，由于逻辑讲求语义明确而严格，但要是日常交流用一堆符号估计谁也受不了，所以科学哲学也在语义学方面继续发展。英国人的经验论也促进了新实验主义与主观贝叶斯学派的发展，慢慢地科学哲学也开始接受一些非实在论的观点，而科学实在论是穿插在上述命题中的。

科学哲学从实证主义发展到今天，被各种新命题与发现折腾的够呛，从里面提一个片段就可以看到很多，科学是什么？它跟哲学啥关系？又对哲学发展有什么样的影响？总之，我们没有停下探索真理的脚步，答案在哪里也毫无头绪，只要不满足于现状，知识就存在进步的可能，同时须知人生苦短，自知无知是很重要的。

就科研本身而言，最开始属于观察现象然后总结规律的经验方式，后来慢慢形成学科体系与知识框架来设计实验预测解释事实，现在其实更多是逻辑与经验的混合来解决科学问题。也就是说，学科知识是基础，但问题总出在前沿也就是知识覆盖不到或部分覆盖的地方，经验论与唯理论的斗争时常出现，单纯看经验或者说观察与实验会推动问题的解决，但有时候也推不动：很多规律不一定经得起检验，还有很多规律需要的限定条件太多进而导致应用上矫枉过正，还有些学科提出规律本身产生的反馈会导致规律失效...

2.2 科研职业化-问题为导向

喝完老鸭汤我们还是回归现实吧，现代科研隶属于现代政治经济系统，满足社会的需求是其存在的基础，至于是否满足个人兴趣爱好与远大理想，可认为是副产品。当然，这是从社会层面说，具体到个人千差万别。

首先，我们要了解现代社会运行的基本模式，其中陌生人大尺度分工协作是现代社会最突出的特色。社会，简单说就是一群人而不是一个人生存的行为与知识模式集合。相比宗族或家庭为单位的原始聚居，古代与近代社会的发展不断突破着人们行为与知识范围的地理与血缘限制。

在原始聚居条件下，人们终生活动范围有限，语言隔阂等也限制了信息交流，好的生存模式很难传递到下一代或更远的地方，短暂的寿命基本都用在维持生存繁衍上了。当然，对原始部落的研究发现生活在其中的人并不比焦虑的现代人的快乐感受更少，但生活的自由度其实很有限（从另一方面讲，如果完全意识不到当今生活自由度可以改变其实也是一种内在幸福感，拥有更大自由度的人并不能完全体会到）。这种狩猎采集的原始聚居其实并不太需要共同的社会行为规则，但后来人们驯化了农作物与牲畜（其实很难讲谁驯化了谁，作物与牲畜也可能通过驯化更好的传播了基因），进而从流动走向了定居。

定居后的社会出现了更细致的分工，例如一个村落需要祭祀、防卫、生产、医疗等部门维持生存结构，这种分工有着自己的生命力，一旦产生会让整体受益，同时也会让这种结构加强。同样的，这种分工模式并

不惟一，但如果两个定居的社会共同体产生利益矛盾，最后剩下来的总是一种更有利群体生存的模式，这个模式下的规则并无道德可言，或者说这就是社会道德的起源。这同时也是一个路径依赖的过程，总会带有一些副产品，很多时候我们就是通过副产品来回溯过去。如同对进化过程的研究一致，使用幸存者就是最好的或最合理的逻辑是不恰当的，我们需要通过回溯来发现一些制度历史上的合理性与偶然性，逻辑自治并不代表历史真相，这点对科研认识也是很重要的。然而，这个阶段的社会政治经济体制依然很大程度被自然条件所控制，多数规则要么偏向农业社会，要么偏向海洋经济，人类的视野逐渐开阔，但基于血缘与地域的多样化依然可以保留，直到更追求效率的技术与体制规则进一步交互作用，孕育出近代工业社会。

近代工业社会将分工与效率推向了极致，影响的范围从多个国家推广到了全球。伴随而来的就是一套基本抛弃血缘关系与多样性的陌生人交流法则，地理限制被信息技术与交通技术打破，所有国家都会遵循同样的工业标准，语言也尽可能一致，法律也会去遵循共通的法则。科学研究在这个过程中起了很重要的作用，而工业化也不断向科研提出需求，此时科学研究从精英们的兴趣爱好变成了巨大的财富来源，每一次技术革新都服务了社会，而几乎所有的社会经济体制都会拿出资金支持科研。务实一点的国家或企业会对工程学优先发展，而对自然科学的支持则颇有情怀意味，毕竟一旦经济下滑，最先拿不到钱的都是基础科研等见效慢的学科。这种社会整体的功利主义自产生之时就展示了巨大的生命力，甚至不断影响了社会中个体的决策行为。

时至今日，现代社会基本延续了工业社会对分工与效率的追求，但维持文化多样性与个体-社会相互关系的思考不断涌现。现代社会塑造了个体认知，个体认知却反过来反思现代社会的诸多问题例如民族主义的崛起、环境保护、气候变化、社会隔离与歧视、机会公平、人口老龄化、战争暴力、谣言传播、经济危机、金融危机、人工智能等。这些问题的根源有相当比例是社会政治经济体制的构建过程出现了漏洞，而今天的科技发展把一些问题放大了，或者说这个系统需要打补丁了。毋庸置疑，科研对于现实问题的解决是一个靠谱的选择，其他选择例如宗教、回归原始生活更多的是一种消极的保守策略，选择那些方法并不会真的解决问题。这样如果给科研立块大牌坊，我想最好的题词就是从方法论层面解决社会问题。换言之，科研总是面向问题解决的一个社会分工，是一个职业，既不神圣也不低俗，从事这个职业的人总在用科学方法论实际问题，有时候也是揭示问题或为问题找一个解释。这个需求是根源，也就是说如果你科研自认为做的不错但跟现实脱节，那么即使留在象牙塔，也会面临自我认同与社会认同不协调的困境，需要你额外的资源平衡。

放到经济视角下，这个职业也是有温饱小康问题的，也是一个利益集团，需要人代表到国会或人大去抢财政的分配，还要跟不同学科去抢所有的科研分配，充满了复杂的博弈过程，原来是陌生人之间，以后可能会发展到人跟机器或规则之间。这个职业有光环，但退却光环都是一个个为生计奔波劳碌的现代人，当然，不同人有着不同的生计标准。

2.3 科研精细化-体面的博士

精细化促进了分工效率限制了个人视野螺丝钉体面的博士通识教育与精英教育

想成为生物医药领域 pi 的研究生博后要在 10 年内搞出 8 篇一作才有戏，而 H 指数预测性很差，业界可拿这个数据到学术圈挖人

2.4 科学知识的五个层次

2.4.1 背景组

高中毕业水平。不论你学的文还是理，知识一般侧重原理或事实本身，或者说学到的是通识，例如地球是圆的、力学有三大定律、元素周期表是按什么排的...这类知识其实就算老师不教，你看看《十万个为什么》什么的也大概能知道。

一般而言，科普主要面向知识背景是高中组的，因为绝大多数人不进行科研，就算进行科研其很多科学背景知识也是高中的，因为你大学可能学了某个学科，但另外的学科最理想也是停留在高中阶段。但这部分知识基本不用科普，或者说包含在更广泛的知识普及中就好了，需要思考推理的部分不多，主要是了解事实，形成背景概念。

2.4.2 已知的已知组

大学毕业水平。主要指专业知识，例如 pm2.5 是怎么回事？行星间距离如何测量？端粒长度跟寿命关系...这个是社会大多数人科学背景知识的上限，也是职业化的下限。如果继续深究，基本就是做这一行的才了解，需要经验。目前这个层次的科学知识几乎可以被维基百科覆盖。如果你本科是理工学科，那么拿到学士学位就表明你已经掌握了这部分内容。这部分知识一般有自己的学科框架跟体系，但学科间壁垒明显，如果知识是构建在经验上的那么很有可能被机器超越而失业。

2.4.3 已知的未知组

这部分的知识是从已知走向未知，知识都是比较前沿的，很可能被后续结果推翻掉。所以这部分内容是需要不断研究的，但就研究思想层次看，其实即使一线科研工作者自己可能都比较迷糊，大部分人是站在前人基础上往前推进，前人的研究结论容易保存，但思想可能早就消散了。这阶段的知识因为开始走向应用所以学科间开始互通。科研人员的知识水平基本是这个层次的，知道前沿在哪里但还在探索。统计学也是这个阶段常用手段，但需要使用者理解方法本质，人工智能可作为研究方法但无法职业替代，因为探索性的思路还得人来启发。

2.4.4 未知的已知组

这个类别文章侧重于整合已有知识进行创新得到的新知识，基本上只有经验很丰富的人才能站到一定高度上理解一个学科。此时谈思想谈逻辑之外要整合学科历史，把发展沿革搞清楚。学科间知识高度互通，只有有洞见的科研人员才能掌握，统计学或人工智能在这个层次知识的失灵，只能摸着石头过河。

2.4.5 未知的未知组

这个组就是天花板了，或者走向科幻了。不过我觉得哪怕是科幻也是要读的，因为你可以体会其中思考的乐趣，当然知识背景设定就不要管了。这个组需要你能构建出一个逻辑自洽且符合现实的知识理论体系，

其实就是形而上的东西了。我不认为现实世界中可以找出这样的人，但小说中是可以虚构出来的。

2.5 知识体系的时间构建

除了基于已知未知这种简单的面向科研划分方法，个人知识还需要一个时间与逻辑上的构建，这样可以自洽于其他非科研知识，毕竟科研是有噪音边界的，但知识却不一定有。

《四库全书》采用了经史子集的划分方法，内在逻辑是把经典的普世价值、过往的经验事实、学科知识与个人文集进行了区分，这样我们可以很容易找到知识的可靠性与归属。个人知识也可以分为形而上的观点理论与形而下的事实经验来区分，日常所见都是事实，前人所见则为历史，幻想与虚构的故事也是未来的存在，所以一个基础的形而下知识体系要有个人经验与历史，侧重对事实的准确描述，而关于未来则可单独开列，因为这类事实并不存在但不妨碍有思考的乐趣。至于形而上的东西，没必要确立经典，要按照逻辑自洽的原则去整理，包括有证据有逻辑的强理论；有逻辑无证据或弱证据的观点及个人经验；如果仅有证据没有观点，可移到形而下部分；仅有逻辑的理论是很危险的，若是科幻小说还值得看，否则不要随意吸收，因为这部分有可能指导改变世界，但尚属于探索，如果你不打算从政或成为企业家，可参考，但如果打算，就要考验个人决策能力了。科学探索恰恰可能也在这个地方，所以得学会用从观察与实验中提炼知识的能力，至于是否改变世界，那是你的自由。

简单说就是你可以创建一个笔记系统来整合你的知识：形而下的基础数据与案例库、历史沿革及信息检索方法与形而上的理论观点库，区分强理论与弱理论并学会总结理论。要学会把新信息整合到这个知识体系中去并持续整合，形成完整独立的知识库，这样就不容易被新思想所迷惑，总能找到位置。

Chapter 3

科研现状概览

3.1 国内版

研究生人数比例研究所：高校 1:3

3.2 国际版

3.3 趋势

3.3.1 科学方法

- 基础学科信噪比高企与泛应用化
- 假设检验的困局
- 反馈对规律的影响：星座
- 平均律的终结
- 实际问题导向的基金流

3.3.2 数据驱动的科研

- 多重检验问题与 p 值

疾病多个阶段都有指示物，根据不同阶段指示物进行综合判断，权重医生来定频率

- 概率是一个过程特性而不是结果
- 我不关心这个数可能是什么分布，只关心这个数具体是什么

- p 值这个问题，重要的不是把 0.05 降到 0.005，通用阈值这个想法太偷懒，应该让研究人员充分理解 p 值实际意义与使用方法，毕竟在有些研究领域控制 fdr 后阈值实际比 0.005 低得多
- #nature# nature 就近年来的科研可重复性危机采访了五组科学家，分别从认知、NHST、FDR、数据共享与范式转化的角度进行了论述，当然别忘了到 Gelman 大人的博客上围观他的花式吐槽

3.3.3 假设检验问题

- 发现的价值不依赖 p 值，依赖效果大小与参数，进一步依赖样本量
- p 值常被认为随机生成的可能性，但混淆了 $P(D|H_0)$ 与 $P(H_0|D)$, p 值给的是前者，要想知道随机生成的概率，需要知道空假设是真的概率
- p 值还被认为结果的可重复性，但重复性是统计功效的函数，跟 p 值无关， p 值不能传达真实与否的信息
- NHST 用来评价空假设为真时观察概率，但多数情况不显著的结果是在研究假设而不是空假设下进行检验
- 统计功效很重要，跟样本数关系大
- 当样本数增大时，空检验总会被拒绝，因此当空假设为感兴趣的理论时，样本数与准确性会提高理论强度，但空假设不存在时，样本数与准确性提高只会弱化理论
- 发表歧视，科研文献不科学反应研究现状
- 基于 p 值还不如描述性统计
- p 值控制只考虑假阳性而不是假阴性
- 低功效研究会实际扩大研究的效果

John Myles White 系列

- 隐藏多重检验，检测多个变量，留下显著的
- NHST 最大的问题在于解决问题的决策者与学术圈的认知不一样，研究人员的使用知识是不全面
- 对于个体而言可能还好，但对学术圈来说有时候是黑盒子
- Cohen 的国会议员例子，空假设是某人是美国人，备择假设是非美国人。我们知道某人是国会议员的概率是百万分之二，空假设里很难发生，备择假设里无法发生。空假设我们拒绝了某人是美国人，那么根据 NHST，他不是美国人。但问题是议员一定要是美国人，在此类问题里，NHST 永远无法认定稀有事件，也就是功效永远不足，并会给出错误答案。稀有事件总会发生，NHST 总会把此类事件当成显著，即使不那么稀有，例如小概率事件如果发生了，我们就可以拒绝了。
- 进行假设检验时，我们是在测定空假设成立时数据发生的概率， $P(D|H_0)$ ，但我们真正关心的是 $P(H_0|D)$ ，也就是数据是否支持假设，但这需要贝叶斯定律来计算
- p 值被用在同时测定效果强度与不确定性上，特别是很小但是很精确的差异，这是测量的进步，但却是理论的噪音，这样需要置信区间
- p 值的应用领域正在从硬领域走向软领域，低垂果实已经没了 Paul Meehl: The almost universal reliance on merely refuting the null hypothesis as the standard method for corroborating substantive theories in the soft areas is... basically unsound.

- NHST 已经不能帮我们积累知识，考虑一个多元线性模型，我们只能在多元模型里得到参数，也就是有限检验，不能发现未知参数，但科学就是寻找未知；变量间的关系在数值改变后如何考察，正负关系如何预测，预测性无法实现
- p 值正在成为测量投资与努力而不是事实的标准，给定差异，我们总能找到足够的样本来发现这个差异，如果打算测量就报告测量，p 值并不能增加知识

3.3.3.1 对策

- 停止使用假设检验，回归理论
- 报告描述性统计，置信区间，效应大小与参数
- 如果用，报道功效
- 学科学哲学与科技史
- Meehl: We should treat the history of science as a scientific problem.
- 学数学: Use mathematical models to understand and explain the phenomena of interest; do not force the phenomena to fit a readily available, generic mathematical model.

贝叶斯

- 概率来自数据，长期表现需要分开讨论
- 我不关心我没有做过的实验，只关心基于当前实验能让我对这个数的估计改变多少
- 置信区间与可信区间 (Confidence interval Credibility intervals) <http://stats.stackexchange.com/questions/2272/whats-the-difference-between-a-confidence-interval-and-a-credible-interval>

不用 p 值不会有 type 1 跟 type 2 错误，但是会有 type s 跟 type m 问题，前者是正负标志，后者是数量级

p 值对个体研究者有意义但对群体有害，因为群体不知道个体研究的细节与尝试过程，例如找 20 个 x 与 y 关系总能发现有相关的，但只报道这个结果会导致其他人无法重复

Bayesian

- 多重比较里，p 值是根据你比较数而不是理论决定的，这导致你的主观臆断决定了结果
- 置信区间也会受到实验意图的影响
- 无限假设下，p 值阈值可以无限小，也就是让所有结果都不再显著
- 设想投硬币跟投票，如果 1000 次有 535 次正面或投给候选人 A，NHST 无法区别，但贝叶斯下前者无偏，后者领先
- 贝叶斯方差分析先假设分布，然后用数据更新分布，后验分布计算出来就同时有点估计跟方差估计，同时多重比较问题也不存在，但随机错误无法避免，此时参数估计方差大也能体现，后续研究可以使用这次的后验数据作为下次先验数据

- 效应估计中 95% HDI 95 高密度区间, 空假设为真也会有 5% 次没有真值
- ROPE 真实等价范围, 如果 HDI 不在区间内, 那么拒绝, 如果覆盖接受
- 效应估计的另一个方法是比较模型, 参数可以来自收敛到更大的分布, 这样可以降低错误, 同时有利于空假设而不是拒绝, 覆盖零不意味拒绝, 有分布支撑
- 存在模型比较后空模型占优但是参数 HDI 排除了空模型, 这个情况由于两个模型的先验概率对结果是敏感的
- 三种重复性: 假设模型仿真看结果; 后验模型仿真看结果; 后验模型仿真后验模型做先验假设看累计效果
- 贝叶斯与频率学派之争

<http://andrewgelman.com/2012/07/31/what-is-a-bayesian/>

<http://andrewgelman.com/2012/02/24/untangling-the-jeffreys-lindley-paradox/>

<http://andrewgelman.com/2014/01/16/22571/>

https://en.wikipedia.org/wiki/Lindley's_paradox#The_lack_of_an_actual_paradox

<https://xianblog.wordpress.com/2014/02/04/posterior-predictive-p-values/> - 可重复性问题

3.3.4 社交网络中的科研

- 多媒体 v.s. 文字
- 快速反馈交流
- 合作全球化与圈子化
- 跨学科研究中的囚徒困境

Chapter 4

思维工具篇

4.1 科学思维

4.1.1 规律的失效

科学靠谱很大程度是规律性进行的保障，规律保证了可预测性，但其实很多规律恰恰说明在很多地方没有规律。

4.1.2 哈森奇效应

4.1.3 观察研究的敌人-反馈

4.2 模型思维

4.2.1 可编程

- 正则表达式
- 递归
- 循环

4.2.2 抽象

4.2.3 交互作用

4.3 统计思维

4.4 估算法

4.4.1 费米估计

<http://www.mathsisfun.com/numbers/estimation.html>

<http://teachersinstitute.yale.edu/nationalcurriculum/units/2008/5/08.05.06.x.html>

Chapter 5

实验

5.1 实验设计原则

5.2 定性实验

5.3 定量实验

5.4 思想实验

- 人是否在仿真中

Chapter 6

数据处理

6.1 多重比较

6.2 多重检验

6.3 回归

6.4 预测

6.5 仿真

6.6 可视化

- 周期性作图需要画两个周期来观察其变化相关
- 生物数据可视化 <https://www.nature.com/articles/nbt.1567>

主题:

- 假设检验与 p 值 q 值与 fdr 多重比较
- 相关性分析与工具变量 <https://www.nature.com/articles/nbt0309-255>
- 线性模型中混杂因素的判定与消除
- 聚类与主成分分析 <https://www.nature.com/articles/nbt0308-303> <https://www.nature.com/articles/nbt1205-1499>
- 列联表分析与流行病学研究

- 多重比较问题 <https://www.nature.com/articles/nbt1209-1135> Bretz, F., Hothorn, T., Westfall, P., 2010. Multiple Comparisons Using R. CRC Press. Gabriel, K.R., 1978. A Simple Method of Multiple Comparisons of Means. J. Am. Stat. Assoc. 73, 724. <https://doi.org/10.2307/2286265> Gelman, A., Hill, J., Yajima, M., 2009. Why we (usually) don't have to worry about multiple comparisons. ArXiv09072478 Stat. Plotting of multiple comparisons? [WWW Document], n.d. URL <http://stackoverflow.com/questions/2286085/plotting-of-multiple-comparisons> (accessed 11.9.13). Rafter, J.A., Abell, M.L., Braselton, J.P., 2002. Multiple comparison methods for means. Siam Rev. 44, 259–278. Stoline, M.R., Ury, H.K., 1979. Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons among Means. Technometrics 21, 87. <https://doi.org/10.2307/1268584> 多重比较谬误, [编辑], n.d. . 维基百科, 自由的百科全书.
- 统计重采样模拟如 bootstrap 思想
- 偏最小二乘分析在结构效应关系研究中的应用
- logistic 回归与剂量效应曲线
- 人工神经网络与黑箱计算 <http://www.nature.com/doifinder/10.1038/nbt1386>
- 从 lasso 到岭回归, 惩罚项在回归分析中的应用
- 截断回归与缺失值处理
- 线性混合模型
- 支持向量机的回归与分类 <https://www.nature.com/articles/nbt1206-1565>
- 从决策树到随机森林 <http://www.nature.com/doifinder/10.1038/nbt0908-1011>
- 线性判别分析与特征发现
- 经验贝叶斯与近似贝叶斯计算 (ABC) 及贝叶斯网络 <https://www.nature.com/articles/nbt0106-51> <https://www.nature.com/articles/nbt0806-959> <https://www.nature.com/articles/nbt0904-1177>
- 时间序列分析
- 结构方程模型
- 分层模型 <https://www.nature.com/articles/nbt.1619>
- EM 算法 <https://www.nature.com/articles/nbt1406>
- 动态规划 <https://www.nature.com/articles/nbt0704-909>
- MCMC 方法 <https://www.nature.com/articles/nbt1004-1315>

Chapter 7

文献

7.1 文献管理

文献管理方面主要包括文献收集、整理、分析与追踪，目的是获取当前研究趋势。用认知过程阶段可以分成三个：从无到有、从有到精与从精到用，从无到有是指刚进入一个新领域时的状态，绝大多数研究生跟转行的科研人员都要通过这个阶段构建自己的文献知识库；从有到精指维护与整理与追踪新文献；从精到用阶段指文献知识库体系直接参与科研过程形成产出的过程。

7.1.1 从无到有

刚开展研究工作的第一步就是背景知识的了解，除非你研究生转行，一般本科阶段的学习应该已经掌握了学科基础，这个是共通的背景知识。基于此你要从教科书上相对确定的知识走向文献资料中相对不那么确定的知识，此时最好的开端是一本英文教材，一方面锻炼英文，另一方面英文教材的更新比国内要快（你大概率可以从图书馆借到，而且多数图书馆都有根据你需求订书的服务，不要浪费）。如果你精力足够，甚至可以联系作者问下是否可以翻译，这样一举多得，不过我没操作过，只是建议。另一个思路是通过 MOOCs 来系统学习，国内外很多高校放到网上的课程授课老师都属于接受新思想比较快的人，讲义也比较前沿，系统性比较高。还有一个不太通用的方法是阅读近些年的博士论文，其文献部分一般都是相关信息，不过能不能找到就不好说了。这个阶段一般要两三个月，不要心急，先把基础打好。前面掉的坑越多，后面跳坑就更有经验。

一般而言，一项学术成果要先发表，然后被综述评论，然后进入研究生课程讨论班，然后进入本科生课程讲义，最后才进入学科经典教材的更新。所以你可以倒着去走这个流程，越往后可能越不容易懂，但循序渐进总比一下读前沿论文被搞晕要好。有了相对前沿的教材或讲义作为知识框架，你的脑子里此时应该比较清楚导师让你做的东西或自己打算做的东西在学科中的定位，解决的是什么科学或工程问题，此时可以进行基于关键词检索的文献收集了。

一个良好的搜索返回的结果应该在 10 篇以内，首先要是综述，然后关键词检索方面建议学点逻辑运算符来过滤掉不相关信息，如果你上一阶段看的书是 5 年前更新的，那就只去关注最近 5 年的综述；如果你

做的领域实在太新，那就把关键词信息的同义词跟近义词也加到搜索里；如果你能找到一篇写的特别好的综述或者有高人指点的论文，那是最高效的方法，可遇不可求。这 10 篇论文请按年为单位每 1-2 年选一篇综述去看，一月内读完，要求是精读，也就是论文里提到的研究都加到你的文献库里并阅读细节，同时可参考综述章节对文献库进行分组。一定要做笔记，而且要进行结构化的笔记或思维导图，这个阶段时间可能比较长也比较累，成果是当你去听系里的报告时，你大概能将报告定位到你的笔记框架里。到此文献库就从无到有了。

7.1.2 从有到精

有了文献库不代表就不用读了，你要建立一个体系来整理并追踪最新文献，这一阶段希望你早就了解 RSS 是怎么回事并且使用过 RSS 阅读器。如果没有，邮件订阅也不失为一个良方。这里我要提示一下，一般文献库管理工具都提供针对单篇文献的笔记功能，不要用。请自建按研究主题的笔记，把新的有意思的新论文连同你以后可能引用的语句直接摘到相关主题的笔记里，而且要让你的笔记可以反链到数据库或通过 doi 可以直接找到原文（推荐后者）。没别的意思，我希望你的笔记稍加整理就可以作为综述发表，省的你次次重返工。建议文献追新频率每周一次，固定时间，看到好的文章就马上消化掉。

7.1.3 从精到用

文献信息的收集与整理不是为了写笔记，是为了需要用的时候瞬间能够用到，例如写一个技术报告，给别人审稿，还有最重要的：写科技论文。科技论文不同于其他文体一个最显著的特点就是参考文献体系的支撑：所有的讨论都要起于前人的发现，参考文献事实上经常是考察作者知识面的关键，对前人工作的遗漏会严重降低文章的系统性与创新性，经常会被审稿人一票否决，哪怕其实你做的跟前人是不同的。另外的使用就是报告幻灯片跟其他学术交流场景，如果你能做到在大脑或笔记中快速定位到一个观点或现象然后几句话说明清楚，这个习惯能帮你离开学术界后在其他行业直接展开降维打击。绝大多数离开学术界的人都不会继续保持了解前沿动态的习惯而更多依赖过往经验，一个人的经验如何去抗衡一堆参考文献背后成百上千人的经验？当然有些东西那些成百上千人也许都不知道，特别是工程上的。不过这种“学院派”的研究习惯最大的好处就是让人更谦逊些，知道一山更比一山高，处处重峦叠嶂。那些上来就趾高气昂且沉醉于自己小圈子的人，不管在学术界还是其他行业，九成以上是鼠目寸光之辈，请远离这些人。

谈文献管理，我希望不要掉到工具选择的坑里，要构建完整的知识管理体系，哪怕是基于便签的只要能实现头脑知识的更新换代就可以了，如果能方便写作投稿，那就更好了。切不可舍本逐末，单纯把文章发表作为目标去优化，毕竟所有的短期目标都要最终整合成你学术生涯的一部分，可以抽时间去想想一些简单的问题：

- 我的研究究竟有没有实际意义？
- 我的发现是否有助于学科发展或写入教科书？
- 我现在纠结的事 10 年 20 年后会不会纠结？

以人之渺小，所有的时间都是浪费，但你要为自己浪费的时光赋值。

7.2 信息收集

- 科技类 nature / pnas / science
- 健康类 NEJM / JAMA
- 综合期刊 plos one/peerj
- 专业顶级期刊
- 会议论文
- twitter 上 #icanhazpdf 文献求助百度小木虫
- rss
- 100% 读题目 20-50% 读摘要 5-10% 看图 1-3% 全文
- 追踪课题组

文献管理软件方面有收费的也有免费的，一般而言可通过咨询自己所在科研机构的图书馆来获取是否购买了相应的软件。收费软件我使用过 Endnote、papers 及国产的 Noteexpress，应该说早期它们之间差异还是明显的，但到今天基本同质化了。

7.2.1 Zotero

7.2.2 Mendeley

7.2.3 EndNote

7.3 文本挖掘

7.3.1 关键词

7.3.2 作者

7.3.3 时空分布

7.3.4 影响力

- 科学家的奖励信号 <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0142537>
- 提前发表可以提高影响因子 <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0053374>
- 题目越短引用越多 <http://rsos.royalsocietypublishing.org/content/2/8/150266>

7.4 荟萃分析

Chapter 8

学术生活

8.1 项目管理

8.1.1 香肠战术与拖延症

- 不断切分到具体可执行
- 执行时不思考整体，关注当下
- 转移注意力可放松，但要设计成杨白劳模式

8.1.2 时间管理

- 进度控制
- 紧急重要四象限
- 与未来自己博弈

8.1.3 笔记管理

8.2 学术出版

8.2.1 期刊论文

- 合作者人数与版本控制
- 先画图后写作
- 可视化陷阱
- 模块化制图

- 语言简洁可能不利于发表，但有利于传播
- 公开代码、数据、软件来提高研究的可重复性，被重复有利于提高学术影响力
- 信任合作者
- 马上动笔，拒绝完美
- 博客文章、软件包在传播上与论文一样重要，甚至更重要
- 论文由方法、数据、结果跟结论组成，先完成前三个
- 阴性结果也是结果，对其他科研人员也有参考价值，也要考虑发表，哪怕是博客发表
- 预先发表
- 投稿到你设想中读者会读到的地方
- 通过社交网络传播你的发现
- 开放获取??? 基金与影响力传播
- 延长论文的半衰期
- 回复审稿人
 - 逐条回复
 - 指明文中修改的位置

8.2.2 会议摘要

8.2.3 专著

- 在线出版
- 允许反馈
- leanpub / gitbook / amazon kindle direct publishing / bookdown
- 按需求出版纸质版 lulu.com

8.2.4 专利

8.2.5 软件

软件就是发表 JOSS

8.3 学术会议

8.3.1 口头报告

- 娱乐而不是教诲
- 写给观众而不是自己
- 用所见即所得方式
- 首页有联系方式
- 大字体
- 有链接
- 对比色
- 图片文字比 1000:1
- 解释图片时先说干什么用的，然后解释坐标，然后解释关键现象
- 解释公式时用文字不要用单一符号，脚标不要太多
- 注意时间

8.3.2 海报报告

8.3.3 听报告

8.4 审稿

- 尽快，否则不做
- 可以进行发表后审稿或公开评论
- 流程是主编确认投稿是否合乎范围，分配给专业副主编，副主编寻找审稿人
- 如果对方改了也不能达到你认为的标准，拒绝而不是大修或小修
- Leek, J.T., Taub, M.A., Pineda, F.J., 2011. Cooperation between Referees and Authors Increases Peer Review Accuracy. PLoS ONE 6, e26895. <https://doi.org/10.1371/journal.pone.0026895>
- Logan, B.E., 2014. I Owe, I Owe, so Off To Review I Go. Environ. Sci. Technol. Lett. <https://doi.org/10.1021/ez5001148>
- Lu, J., Law, N., 2012. Online peer assessment: effects of cognitive and affective feedback. Instr Sci 40, 257–275. <https://doi.org/10.1007/s11251-011-9177-2>

8.5 学术合作

8.5.1 数据共享

8.5.2 社交网络

- 记录学科内的重要进展
- 如果个人忙不过来，可以考虑多人合作编辑或找到组织成为作者
- 跟业界联系的渠道
- 提高自己交流与可视化的能力
- 每个人每天都会在互联网上花费时间，博客是有机融合
- 互联网只关注异常、争论与胜负而不是共识，共识可以交给科普
- 质疑要比实际操作轻松
- 对自己文章要按互联网信息传播速度回复或形成论文发表来回应
- 不回复质疑会影响学术声誉
- 博客行文会影响别人对你的看法
- 构建 / 加入在线学术社团
- 放大影响力
- 构建趋势感
- 线上线下互联
- 多使用图片
- 谨慎介入糊涂账话题
- 均值、中位数等单一数值常在媒体报道跟论文中用来指代群体，但其实牺牲了很多重要的分布细节进而产生误导，甚至让人产生被平均的感受，而直接展示整体其实并不困难，重要的是作者/研究者应放开心态，从引导读者认同自己观点转为让读者自己探索出结论

8.6 讲课

- 教案与视频在线化
- 在问答社区贡献答案 quora stackoverflow reddit 知乎果壳
- slideshare speakerdeck figshare
- mooc 课程尽量短（少于 10 分钟）脸皮要厚注意产权

8.7 课题组管理

- slack / hipchat
- 按项目交流
- 定期交流
- 人员管理要及时

- 分享文献
- 分享实验室内部规定

8.7.1 科研的创业隐喻

其实科研，特别是理工科科研，很像是创业过程。互动的理解这个过程有助于青年科研人员知道自己究竟在干什么。

- 投资人：政府或企业
- 公司：项目或课题
- 董事长：课题组长（PI）
- 经理：小老板或子课题负责人
- 项目经理：博士生 / 硕士生 / 本科生
- 员工：无
- idea / 论文：产品
- 质检员：张全蛋，额，打错了，是审稿人
- 产品上市：论文发表
- 产品发布平台：学术期刊
- 产品未通过内部质检：论文拒稿
- 产品有销量：论文被引用
- 产品成为爆款：论文被大量引用
- 产品被媒体推荐：论文被编辑或综述点评
- 产品滞销：论文无引用
- 产品原料配料表：论文数据共享
- 产品生产流程：数据处理
- 产品被模仿：论文被重复验证
- 产品补丁：论文修正
- 产品更新换代：论文跟进发表
- 产品推介会：学术会议
- 产品退市：论文撤稿

- 公司融资：申请基金
- 新三板 / 创业板：申请青基
- A 股：申请面上
- 路演：项目申请书
- 估值：学术影响力
- 竞争公司：研究方向

8.7.2 基金申请

- 基金透明度问题 <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002333>

8.8 学术声誉

- 论文第一
- 定量化
- altmetric
- cv 优化
- 跟老司机交流
- 职业规划
- 常任轨研究教授 / 常任轨文理学院教授 / 研究性教职 / 业界
- 常用 ID 简短难重复
- 商用个人用 email 分开
- 线上线下一致

8.9 学术道德 / 伦理

8.10 案例

- 科学新闻探索
- 华盛顿邮报根据一项研究成果写了篇科学报道，然后捅了统计学家的马蜂窝，Gelman 称之为 “beauty”，因为这项研究基本把实验设计中常见的错误犯了个遍，而作为普利策奖得主的记者完全没看出来
- 学科是存在鄙视链的，例如这位物理学家看到皇家化学学会发表一篇烤面包公式的论文后不但马上发明了一个 bullshit 公式，还声称如果物理学会敢发表这种东西，自己马上放弃会员
- 巫毒娃娃的 PNAS

附录：现代科研兵刃谱

工欲善其事，必先利其器。今天绝大多数知识都是工具生产出来的，也就是想使用知识，肯定要先学工具，而工具又需要知识铺垫，这就成了一个鸡生蛋蛋生鸡的问题。虽然事后总结都有千般道理，但就我人经验而言，工具与知识是相辅相成缺一不可的，过于关注知识会导致脱离实际而沉迷于工具选择则有很高的迁移成本。这里唯一的忠告就是不要想太多，先迈开步子，随便找个工具用起来，用实战来丰富需求，根据需求定向选择最适合自己的工具而不做工具的奴隶，如有必要，自己创造工具。

文本编辑

- 所见即所得
 - 在线协作：google docs / 石墨 / 腾讯文档 / Github
 - 离线协作：word
- 排版/文献管理软件
 - Overleaf
 - sharelatex
 - paperpile
 - crossref
 - zotero

数据处理与绘图

- 所见即所得
 - 百度脑图
 - autodraw 简笔画
 - plotly
 - 自动制图

- 编程绘图
 - R base
 - ggplot
 - python matplotlib
 - matlab
- tidyverse
- shiny

学术交流

- 预先发表服务
 - arxiv
 - biorxiv
 - peerj
 - 接受预先发表期刊的列表
- 学术博客
 - blogdown
 - wordpress
- 幻灯片
 - PPT
 - xaringan
- 学术出版
 - bookdown
 - rticle
- 网络身份
 - ResearchGate
 - ORCID
 - ResearcherID
 - Google Scholar
 - Scopus Author ID

审稿

- Publons 记录审稿记录
- 博客或微博审稿
- 期刊网站评论
- Pubmed Commons
- f1000research 上付费发表

数据分享

- figshare
- Open Science Framework
- Dataverse
- zenodo
- 包含原始数据，处理后数据，数据收集的信息与处理代码
- 随论文一同公布
- 尊重数据生产者

代码管理

- 公开代码并记录版本 github 或 bitbucket
- 链接到对应的论文上
- 代码最好经过审查 cran bioconductor ropenscience pypi
- rmarkdown jupyter 文学化编程
- 给未来的自己做注释
- 记录运行环境保证重复性
- Good writers borrow from other authors, great authors steal outright
- Docker image

R 包管理

- Rstudio 小抄
- Rstudio package 模版
- 格式化代码 formateR Rd2roxugen
- 写文档 roxygen2
- 写小品文 rmarkdown
- 单元测试 testthat
- 代码执行效率可视化 profvis

- 集成在线测试 `travis-ci`(linux) `appveyor`(windows)
- R 包网站 `pkgdown`
- R 包教程 `learnr`