

doi: 10.19920/j.cnki.jmsc.2024.03.001

P 值之争与管理学研究: 先验概率的意义^①

贾良定¹, 林泽民², 王涛³, 卜濛濛⁴, 何刚⁵

(1. 南京大学商学院, 南京 210093; 2. 美国得克萨斯大学奥斯汀分校政府管理系、统计与数据科学系, 奥斯汀 78712; 3. 法国里昂商学院战略与组织系, 埃居利 69130; 4. 南京财经大学工商管理学院, 南京 210023; 5. 北京大学光华管理学院, 北京 100871)

摘要: P 值之争的本质是人们把统计检验和统计推断混淆, 直接用统计检验结果来进行统计推断. 这导致许多似是而非的理论被接受. 统计检验犯错的是假阳性概率. 统计推断犯错的是假阳性之反概率. 两种概率之间的数值关系受先验概率影响很大. 先验概率越高, 利用统计检验的结果进行推断所犯错误概率就越小. 如果先验概率很小, 即使假阳性概率很小, 利用该结果进行推断所犯错误概率也很大. 本文以管理学研究为例, 提出解决该问题的办法是, 通过与现实和与理论的双重对话来掌握先验概率, 以提高基于统计检验的结果进行统计推断的可靠性. 本文的科学哲学意义是, 揭示了科学研究是先验与经验相统一的本质, 以及科学研究的不充分决定性的特征.

关键词: P 值; 零假设检验; 假阳性之反概率; 先验概率; 管理学研究

中图分类号: C8 文献标识码: A 文章编号: 1007-9807(2024)03-0001-14

0 引言

“ P 值 (P values)”可能是现代科学研究中使用最普遍的语言了. 研究者普遍使用 P 值进行假设检验和统计推断, 使用最广泛的方法是零假设检验 (null hypothesis significance testing, NHST)^[1]. 在现代科学中, 研究者对基于 P 值的 NHST 方法的信奉达到了“宗教信仰地位”^[2]. 但是, 人们对基于 P 值的 NHST 方法质疑不断, 特别是最近几年, 批评激烈. 批评主要是三方面:

一是错误理解和使用. 如 Goodman^[3] 指出对 P 值的 12 种错误理解, Greenland 等^[4] 指出对 P 值、置信区间和统计功效的 25 种错误理解. 2019 年《自然》发表一篇来自全球 50 多个国家或地区的 800 多位学者 (包括统计学、生物学、心理学、临床学和医学等) 签名支持的评论文章^[5],

同时《美国统计学家》发表专刊^[6], 要求彻底放弃两分法的“统计显著性”错误说法和用法.

二是该方法的不足之处. 主要有: 1) P 值不能进行直接推断. P 值是关于数据 D 在 H_0 为真下的条件概率, 即 $\Pr(D|H_0)$, 但需要推断的是, 在数据 D 下 H_0 为真的概念, 即 $\Pr(H_0|D)$. 这两个条件概率并不相等, 所以不能根据 P 值直接进行推断^[2,7,8]. 2) 无法为零假设的成立提供证据. 当 P 值大于设定的 α 时, 由于不满足小概率反证法的条件, 其不能直接接受零假设为真^[8,9]. 3) 不能进行多个理论间的比较, 特别是零假设与研究假设之间的比较^[7,10]. 比如说, 如果得到 $P = 0.01$, 是否就拒绝 H_0 而接受 H_1 呢? 不一定. 进一步的推断还要看 $\Pr(D|H_1)$, 即数据 D 在 H_1 下的条件概率.

三是使用该方法的不道德行为. 最典型的是

① 收稿日期: 2019-06-21; 修订日期: 2023-01-20.

基金项目: 国家自然科学基金资助项目 (71632005; 71732002; 72132003).

作者简介: 贾良定 (1968—), 男, 安徽合肥人, 博士, 教授, 博士生导师. Email: jldylzs@nju.edu.cn

P 值操纵 (P -hacking) 和摘樱桃 (cherry picking) 问题. 如《科学》杂志 2015 年报道了一项复制性研究的发现, 他们复制了 100 个心理学实验, 虽然 97% 的原研究发现了显著性效果 ($P < 0.05$), 但仅 36% 的复制性研究发现显著性结果^[11].

针对这些批评, 为了提高研究结果的可复制性和产生可靠的知识, 有学者如 Johnson^[12] 以及由 Benjamin 等^[13] 领衔的 72 位学者提出, 把统计显著性水平由原来的 0.05 降低到 0.005; 有学者提出, 放弃 NHST 方式, 禁止统计显著性检验 [5, 14]. 还有学者提出, NHST 方法以及内含其中的 P 值是清白的, 不要一味地摒弃基于 P 值的 NHST 方法, 而是与其他方法互补, 引入这些方法^[15-17].

本文认为 P 值之争揭示的是, 对基于 P 值的 NHST 方法的误解和错用, 即没有区分统计检

验和统计推断, 用统计检验结果直接进行统计推断, 从而导致错误的统计推断. 解决的办法是, 不是放弃基于 P 值的 NHST 方法, 而是引入先验概率. 先验概率是指, 要验证的假设及其背后的理论在抽样数据之外所知道的发生概率. 研究者通过与现实和理论的双重对话来掌握先验概率, 提升基于统计检验结果进行统计推断的可靠性.

1 科学探究过程中的推理、检验和推断

科学探究是“问题 - 理论 - 证据”三位一体的过程^[18, 19], 包括推理 (reasoning)、检验 (testing) 和推断 (inference) 三大活动, 如图 1 所示.

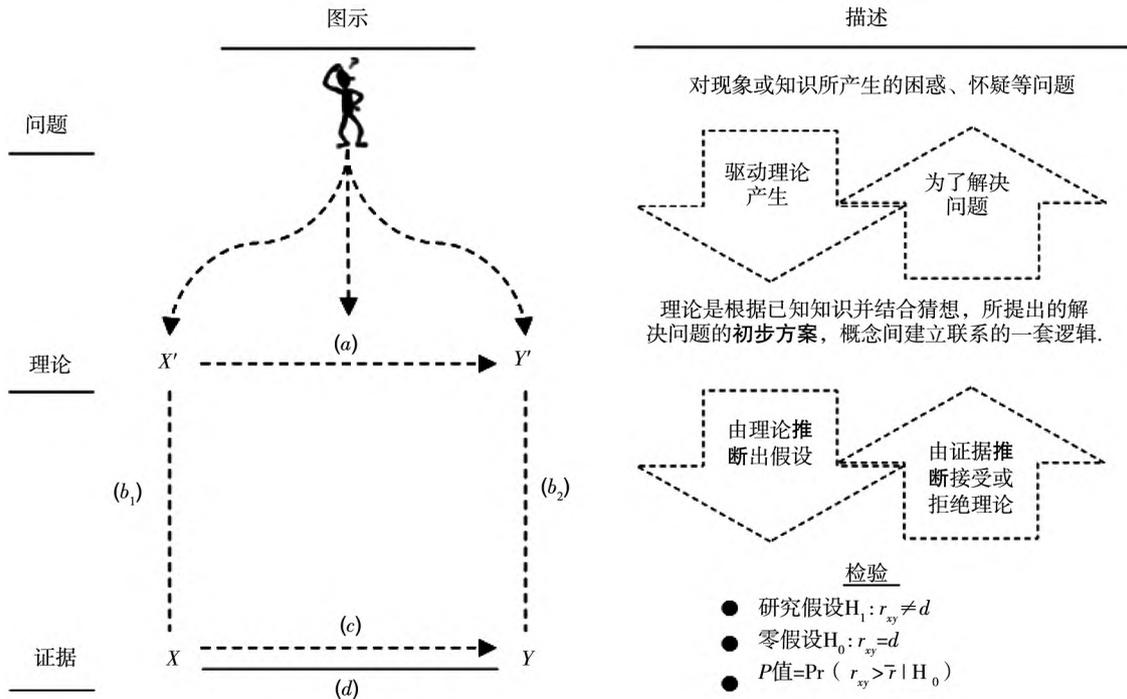


图 1 “问题 - 理论 - 证据”三位一体的科学探究过程

Fig. 1 “problem-theory-evidence” trinity process of scientific inquiry

注 1: X 和 Y 是理论层的概念, X 和 Y 是证据层的变量. X, Y 以及两者间实证关系 (d) 是可以观测到的. 虚线 (a) 是概念间关系, 虚线 (b_1)、(b_2) 是构念的信度和效度, 虚线 (c) 是内部效度. 虚线所表示的内涵都是观测不到的, 只能根据观测到的结果来推断.

注 2: 该图整体结构启发于 Dewey^[18] 的“审慎性思维 (reflective thinking)”以及 Kerlinger 和 Lee^[19]、Schwab^[25].

1.1 推理

科学探究是问题导向的, 问题来源于人们对现象或知识产生的困惑与怀疑^[20, 21]. 根据已掌握的知识并结合猜想, 人们提出解决问题的初步方案. 初步方案是待检验的理论. 人们必须寻找足够

的证据, 直到支持所提出的理论, 才暂时接受^[18]. 理论是概念间建立联系的一套逻辑^[22]. 理论概念间逻辑关系不可以直接在具体情境中进行观察和检验, 但可以通过推理, 形成一系列可以在具体情境中观测和检验的研究假设 (H_1).

1.2 检验

人们能够观测到的是现象世界中的 X 和 Y , 以及据此数据计算出来的两者间的关系 (d). 逻辑上说, 证实“ X 和 Y 间关系是 d ”需要无限次的抽样观测, 而且每次得到的都是 d ^[19]. 但是, 如果“ X 和 Y 间关系是 d ”, 那么有限次抽样得到两者关系为超出 d 一定范围是小概率事件. 因此, 如果一次抽样出现了这一小概率事件, 那么逻辑上说, 每次抽样都能够给出反驳“ X 和 Y 间关系是 d ”的事实.

零假设 (null hypothesis) 必须是准确地与某一值相等的假设^[23]. 假设“ X 和 Y 间是一确定关系 d ”是零假设. 根据有限次抽样观测, 不能直接证实零假设, 但可以提供证伪零假设的事实. 但是如果零假设被证伪了, 并不代表研究者便可以接受在逻辑上与零假设对立的命题. Fisher^[23] 并未提出备择假设这个概念, 提出备择假设这个概念并把它与零假设并立检验的是 Neyman 和 Pearson^[24].

在零假设检验中 P 值的含义是, 如果零假设为真的话, 本文在抽样数据中观察到的该结果或更极端结果的概率. 通常设定一个阈值, 如 $\alpha = 0.05$. 如果 $P < \alpha$, 则 H_0 在该抽样数据的检验中被拒绝. 一般把 α 称为“第一类错误”或“假阳性概率” (probability of false positives), 含义是, 如果 H_0 为真, 统计检验却把 H_0 拒绝了, 所犯错误的概率, 即 $\Pr(H_0 \text{ 检验被拒绝} | H_0 \text{ 为真}) = \Pr(P < \alpha | H_0 \text{ 为真}) = \alpha$. 这是检验过程.

Neyman 和 Pearson^[24] 提出的一个引理奠定了今天把备择假设与零假设并立检验的理论基础. Neyman 和 Pearson^[24] 首先引进了“检验强度”的概念: 当备择假设为真时, 统计检验能够拒绝零假设的概率. Neyman 和 Pearson^[24] 的引理界定了在固定的 α 之下让检验强度为最大的条件. 如果用 β 代表备择假设为真时统计检验未能拒绝零假设的概率 (即第二类错误), 那么检验强度等于 $1 - \beta$. β 也称为“假阴性概率 (probability of false negatives)”, 即如果 H_1 为真, 统计检验却未能把

H_0 拒绝, 所犯错误的概率.

1.3 推断

推断分两步, 先是统计推断. 根据有限次抽样数据计算, 通过证伪零假设 (即 $H_0: r_{xy} = d$), 从而推断, 接受备择假设 (即 $H_1: r_{xy} \neq d$). 然后是理论推断, 推断概念 X 和 Y 间的逻辑关系 (a) 为真.

统计推断的本质是, 基于统计检验, 推断 H_0 为假, 从而接受 H_1 为真. 统计推断要回答的问题是: 如果统计检验时根据给定的标准 (比如 $P < 0.05$), H_0 检验被拒绝, 那么在此条件之下, “零假设为真”的概率是多少? 即 $\Pr(H_0 \text{ 为真} | H_0 \text{ 检验被拒绝}) = ?$ 这个条件概率和假阳性概率的前件和后件是反过来的, 所以称作“假阳性之反概率 (inverse probability of false positives)”. 这两个概率原则上不相等; 只有在 $\alpha = 0$ 的时候^② 或 $\alpha = \beta$ 且 $\Pr(H_0 \text{ 为真}) = \Pr(H_1 \text{ 为真}) = 1/2$ 的时候^③, 两者才相等.

根据贝叶斯定理, 计算“假阳性之反概率”除了需要假阳性概率 (α) 和假阴性概率 (β), 还需要先验概率 (prior probability) 这一信息. 先验概率, 即公式中的 $\Pr(H_1 \text{ 为真})$, 是指在没有统计检验之前, 根据各种已有信息 (如报道、报告、历史资料等) 估计假设 H_1 为真的概率.

假阳性之反概率 = $\Pr(H_0 \text{ 为真} | H_0 \text{ 检验被拒绝}) =$

$$\frac{(1 - \Pr(H_1 \text{ 为真})) \times \alpha}{(1 - \Pr(H_1 \text{ 为真})) \times \alpha + \Pr(H_1 \text{ 为真}) \times (1 - \beta)}$$

必须指出的是, 统计推断只是推断过程的要件之一. 如图 1 所示, 做出可靠的推断还必须满足至少如下两个条件: 一是对概念 X 和概念 Y 的测量, 其信度和效度必须达到要求 (虚线 b_1 和虚线 b_2 表示); 二是满足 X 和 Y 间统计关系的内部效度 (虚线 c 表示).

2 先验概率在统计推断中的作用

根据上述“假阳性之反概率”计算公式, 进

② $\alpha = 0$ 意味着检验是确定性事件, “如果 A 则 B”与“如果非 B 则非 A”是完全等价的. 但是, 如果 $\alpha \neq 0$, 那么“如果 A 则 B”与“如果非 B 则非 A”就不是完全等价的.

③ $\Pr(H_0 \text{ 为真})$ 和 $\Pr(H_1 \text{ 为真})$ 分别是零假设和备择假设的先验概率, 当两者都是 0.5 时, 先验概率没有提供额外信息来修正后验概率; 此时若假阳性概率 (α) 和假阴性概率 (β) 相等, 则“如果 A 则 B”与“如果非 B 则非 A”是完全等价的.

行数据模拟 得到表 1. 由表 1 可知 ,Pr(H_1 为真) 这一先验概率对于计算假阳性之反概率影响很大 ,并且比假阴性(β) 的影响大得多. 当 $\beta = 0.05$ 时 ,要保证 Pr(H_0 为真 | H_0 检验被拒绝) 小于 Pr(H_0 检验被拒绝 | H_0 为真) ,Pr(H_1 为真) 这一先验概率必须不小于 0.5; 当 $\beta = 0.25$ 时 ,要保证 Pr(H_0 为真 | H_0 检验被拒绝) 小于 Pr(H_0 检验被拒绝 | H_0 为真) ,Pr(H_1 为真) 这一先验概率必须不小于 0.6. 如果 Pr(H_1 为真)

这一先验概率很低 比如表 1 中的 0.01 一行 ,那么 ,即使通过显著性检验($P < \alpha = 0.05$) 并且 $\beta = 0.05$,犯错误接受 H_1 为真的概率不是通常所认为的低于 0.05 ,而是高达 0.981.

“假阳性之反概率”受 Pr(H_1 为真) 这一先验概率的影响很大. “假阳性之反概率”越小 ,做出“拒绝零假设(H_0) 接受统计假设(H_1)”推断就越可靠. 下面举三个例子 ,说明先验概率对做出可靠推断的影响.

表 1 假阳性之反概率

Table 1 Inverse probability of false positives

先验概率 Pr(H_1 为真)	$\alpha = 0.01$ $\beta = 0.05$	$\alpha = 0.01$ $\beta = 0.25$	$\alpha = 0.01$ $\beta = 0.50$	$\alpha = 0.05$ $\beta = 0.05$	$\alpha = 0.05$ $\beta = 0.25$	$\alpha = 0.05$ $\beta = 0.50$	$\alpha = 0.10$ $\beta = 0.05$	$\alpha = 0.10$ $\beta = 0.25$	$\alpha = 0.10$ $\beta = 0.50$
0.001	0.913	0.930	0.952	0.981	0.985	0.990	0.991	0.993	0.995
0.01	0.510	0.569	0.664	0.839	0.868	0.908	0.912	0.930	0.952
0.05	0.167	0.202	0.275	0.500	0.559	0.655	0.667	0.717	0.792
0.10	0.087	0.107	0.153	0.321	0.375	0.474	0.486	0.545	0.643
0.15	0.056	0.070	0.102	0.230	0.274	0.362	0.374	0.430	0.531
0.20	0.040	0.051	0.074	0.174	0.211	0.286	0.296	0.348	0.444
0.25	0.031	0.038	0.057	0.136	0.167	0.231	0.240	0.286	0.375
0.30	0.024	0.030	0.045	0.109	0.135	0.189	0.197	0.237	0.318
0.35	0.019	0.024	0.036	0.089	0.110	0.157	0.164	0.198	0.271
0.40	0.016	0.020	0.029	0.073	0.091	0.130	0.136	0.167	0.231
0.45	0.013	0.016	0.024	0.060	0.075	0.109	0.114	0.140	0.196
0.50	0.010	0.013	0.020	0.050	0.063	0.091	0.095	0.118	0.167
0.55	0.009	0.011	0.016	0.041	0.052	0.076	0.079	0.098	0.141
0.60	0.007	0.009	0.013	0.034	0.043	0.063	0.066	0.082	0.118
0.65	0.006	0.007	0.011	0.028	0.035	0.051	0.054	0.067	0.097
0.70	0.004	0.006	0.008	0.022	0.028	0.041	0.043	0.054	0.079
0.75	0.003	0.004	0.007	0.017	0.022	0.032	0.034	0.043	0.063
0.80	0.003	0.003	0.005	0.013	0.016	0.024	0.026	0.032	0.048
0.85	0.002	0.002	0.004	0.009	0.012	0.017	0.018	0.023	0.034
0.90	0.001	0.001	0.002	0.006	0.007	0.011	0.012	0.015	0.022
0.95	0.001	0.001	0.001	0.003	0.003	0.005	0.006	0.007	0.010
1.00	0	0	0	0	0	0	0	0	0

注: 假阳性之反概率 = Pr(H_0 为真 | H_0 检验被拒绝) ,即给定统计检验拒绝了 H_0 的条件下 H_0 为真的概率.

2.1 例 1: 丹尼尔·卡尼曼著作《思考,快与慢》中的出租车肇事逃逸^[26]

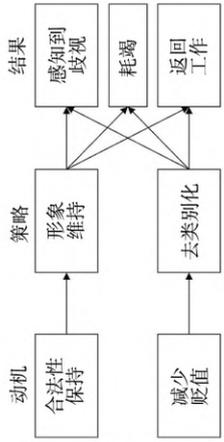
一辆出租车在夜晚肇事逃逸. 该城市有两家出租车公司 ,出租车分别是绿色和蓝色分别占 85% 和 15%. 一位目击证人说 ,他看到肇事出租车是蓝色的. 当晚 ,警察在出事地点对证人的证词进行了测试 ,结论是: 目击者在当时能够正确辨认出这两种颜色的概率是 80% ,错误的概率是

20%. 请问 ,警察该如何裁决? 具体地 ,肇事出租车是蓝色而不是绿色的概率是多少?

这个例子中 ,检验是目击证人所做出的 ,Pr(指认为蓝色 | 肇事车是蓝色) = 0.8 ,Pr(指认为蓝色 | 肇事车是绿色) = 0.2. 推断是警察所要做出的 ,当证人指认为蓝色时 ,肇事车是蓝色的概率是多少 ,即 Pr(肇事车是蓝色 | 指认为蓝色) = ? 下面来分析几种情况: 其一 ,如果没有目击证人 ,

表2 掌握先验概率率的研究实践：以9篇Academy of Management Journal的年度最佳论文为例

Table 2 The research practice of mastering a prior probability: Taking the 9 annual best papers of the Academy of Management Journal as examples

代表论文	研究问题	所构建的理论	与现实对话：基于现实的证据	掌握先验概率率的研究实践：双重对话
Little等 ^[28]	<p>● 为何以及如何进行职业印象管理？</p>		<p>预研究：访谈</p> <ul style="list-style-type: none"> ● 对35名职场怀孕女性的开放性访谈 ● 一方面验证理论推导出动机和策略类别 ● 另一方面区分出不同动机和策略的具体表现，作为发展量表的基础 	<p>整合以下两理论，发展了“职业形象维持理论”。</p> <ul style="list-style-type: none"> ● 职业形象建构理论：个体理想的职业形象和其感知到的职业形象之间的差异会导致个体感知到合法性威胁和贬值威胁，进而分别采取强调社会身份积极的一面与隐藏身份消极的一面的不同策略。 ● 自我调节理论：个体的首要任务是保持原有的职业形象，因此无论个体感知到的是合法性威胁还是贬值威胁，都会采取去类别化和形象维持的策略。
Chua ^[29]	<p>● 周边文化不和谐对多元文化创造力的影响？</p>		<p>实验设计：</p> <ul style="list-style-type: none"> ● 研究二，2×2实验设计，发现只有当被试处于跨文化冲突的情境时，他们文化不兼容的感知才会增强，从而削弱其多元文化的创造力。 ● 研究三，实验室实验设计，排除了消极情感和认知两个替代性解释机制。 	<p>基于内隐信念理论，建立了“多元文化创造力的周边文化不和谐理论”。</p> <ul style="list-style-type: none"> ● 内隐信念理论：当个体在其社会环境中经历史文化不和谐时，文化不相容信念就会被激活或增强。因此，他们把来自不同文化的想法联系起来的能力被削弱了，降低了多元文化的创造力。
Detert和Edmondson ^[30]	<p>● 为什么员工有想法却不愿说出来？</p>		<p>研究一：一家公司深度调研</p> <ul style="list-style-type: none"> ● 在预研中，超过50%的受访者表达过“在工作中对上级进行建言”是令人不舒服的。 ● 大规模问卷调查，大部分回答者对“内隐建言理论”的条目表示认同，认为这些条目反映了他们为何不愿意对上级建言。 	<p>基于以下两理论，建立了“内隐建言理论”。</p> <ul style="list-style-type: none"> ● 内隐理论：存在一些人下意识认为“理所应当”的信念，人们在日常决策中会不自觉地使用这些内隐的信念为自己的决策提供支持。 ● 对权威的认知理论：人们理所应当认为对上级建言会给自己带来危险。

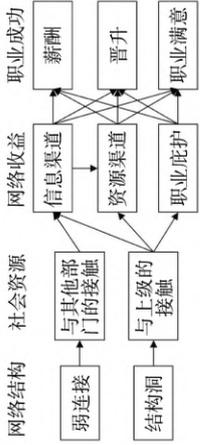
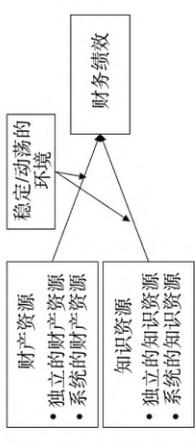
续表 2

Table 2 Continues

		掌握先验概率的研究实践: 双重对话	
		与现实对话: 基于现实的证据	与理论对话: 基于理论的逻辑
代表 论文	Barnett 和 King ^[31]	<p>所构建的理论</p> <pre> graph TD A[自我监管] --> C[行业中其他企业 累计异常收益] B[行业中某一企业的错误] --> C </pre>	<p>基于物质公地的自我监管理论, 发展出“现代行业非物质公地的自我监管理论”。</p> <ul style="list-style-type: none"> ● 自我监管理论: 传统行业内组织由于共享水、土地等物质资源, 进行自我监管, 保护公共资源。 ● 现代行业内的组织, 由于共享非物质的公共资源, 如共同制裁等, 进行自我监管, 保护行业发展。
研究问题	<p>行业中为何会出现自我监管机构?</p>	<p>观察、报道和历史资料:</p> <ul style="list-style-type: none"> ● 作者观察到在某些高科技行业, 衍生企业(由现有公司的雇员出来创办的企业)不仅数量众多, 而且是行业主要创新者。 ● 研究样本是 1977 年—1997 年间磁盘驱动行业, 衍生企业占比很大。 	<p>基于知识基础观, 建立了“衍生公司的知识继承理论”。</p> <ul style="list-style-type: none"> ● 知识基础观: 母公司存在两种至关重要的知识, 技术知识和市场知识, 这两种知识在母公司的不平衡发展, 提高了衍生企业产生的可能性; 平衡发展降低衍生企业产生的可能性。知识继承使得衍生公司的存活率高于其他类型的市场新进入者。
代表 论文	Agarwal 等 ^[32]	<p>所构建的理论</p> <pre> graph TD A[母公司的 • 技术知识 • 市场知识] --> B[衍生公司的产生 • 诞生] A --> C[衍生公司的发展 • 技术知识 • 市场知识] A --> D[衍生公司的绩效 • 存活率] </pre>	<p>基于以下理论, 发展出“制度变革理论”。</p> <ul style="list-style-type: none"> ● 资源依赖理论: 被广泛采用的行业规范使行业内组织对某种资源需求增大, 导致资源稀缺, 资源短缺最严重的事务所更容易成为最初变革者。 ● 制度理论: 较高程度的合法性使某些组织进行早期变革更容易被其他组织所接受。因此在行业中处于较高地位的律师事务所更容易成为最初变革者。
研究问题	<p>什么因素影响了衍生企业的产生、发展与绩效?</p>	<p>历史资料:</p> <ul style="list-style-type: none"> ● 对多位律师事务所合伙人的访谈, 人才短缺是美国律师行业在 1970 年代面临的重大问题, 导致制度变革。 ● 查阅当时新闻报道(纽约时报等), 一家具有声望的律师事务所 Davis Polk 首先推行高级律师轨道, 这一新制度被行业内广泛接受。 	<p>历史资料:</p> <ul style="list-style-type: none"> ● 对多位律师事务所合伙人的访谈, 人才短缺是美国律师行业在 1970 年代面临的重大问题, 导致制度变革。 ● 查阅当时新闻报道(纽约时报等), 一家具有声望的律师事务所 Davis Polk 首先推行高级律师轨道, 这一新制度被行业内广泛接受。
代表 论文	Sherer 和 Lee ^[33]	<p>所构建的理论</p> <pre> graph TD A[行业规范] --> B[制度创新] C[企业声望] --> B B --> D[企业声望] B --> E[制度模仿] </pre>	<p>基于以下理论, 发展出“制度变革理论”。</p> <ul style="list-style-type: none"> ● 资源依赖理论: 被广泛采用的行业规范使行业内组织对某种资源需求增大, 导致资源稀缺, 资源短缺最严重的事务所更容易成为最初变革者。 ● 制度理论: 较高程度的合法性使某些组织进行早期变革更容易被其他组织所接受。因此在行业中处于较高地位的律师事务所更容易成为最初变革者。
研究问题	<p>什么因素导致了制度变革?</p>	<p>历史资料:</p> <ul style="list-style-type: none"> ● 对多位律师事务所合伙人的访谈, 人才短缺是美国律师行业在 1970 年代面临的重大问题, 导致制度变革。 ● 查阅当时新闻报道(纽约时报等), 一家具有声望的律师事务所 Davis Polk 首先推行高级律师轨道, 这一新制度被行业内广泛接受。 	<p>历史资料:</p> <ul style="list-style-type: none"> ● 对多位律师事务所合伙人的访谈, 人才短缺是美国律师行业在 1970 年代面临的重大问题, 导致制度变革。 ● 查阅当时新闻报道(纽约时报等), 一家具有声望的律师事务所 Davis Polk 首先推行高级律师轨道, 这一新制度被行业内广泛接受。

续表 2

Table 2 Continues

代表论文	研究问题	所构建的理论	与现实对话：基于现实的证据	掌握先验概率的研究实践：双重对话与理论对话：基于理论的逻辑
Seiber 等 ^[34]	<ul style="list-style-type: none"> ● 个体的社会资本如何促进个体的职业成功? 	 <p>网络结构：弱连接、结构洞</p> <p>社会资源：与其他部门的接触、与上级的接触</p> <p>网络收益：信息渠道、资源渠道、职业庇护</p> <p>职业成功：薪酬、晋升、职业满意</p>	<p>与理论对话：基于现实的证据</p> <ul style="list-style-type: none"> ● 文中未发现 	<p>与理论对话：基于理论的逻辑</p> <p>整合以下三理论,构建了“职业成功的社会资本理论”。</p> <ul style="list-style-type: none"> ● 弱连接往往是紧密联系的社会小团体之间的桥梁,从而提供了独特的信息和资源。 ● 充满结构洞的网络为个人提供独特及时的信息,更大的议价能力、资源控制和可见性,更多的机会。 ● 社会资本理论关注网络中资源的本质。
Tsu 等 ^[35]	<ul style="list-style-type: none"> ● 不同类型的雇佣关系下,员工的绩效和态度相同吗? 	 <p>期望的贡献：高、低</p> <p>提供的诱因：高、低</p> <p>高：相互投资型、过度投资型</p> <p>低：投资不足型、准现货契约型</p>	<p>新闻报道：</p> <ul style="list-style-type: none"> ● 观察到 1988 年到 1990 年时间的美国企业的一些改变,作者成立讨论小组,每周花两、三小时讨论问题,既看文献,又看报纸,了解实际情况。 ● 针对每种雇佣关系的类型,给出实际的案例,例如准现货合同类型,其经典例子是经纪公司和股票经纪人之间的雇佣关系。 	<p>基于社会交换理论,从雇主角度建立了“员工组织间的雇佣关系理论”。</p> <ul style="list-style-type: none"> ● 诱因 - 贡献双因素:雇主考虑雇佣关系决策时,主要考虑:一是提供给员工的诱因,另一是期望员工的贡献。 ● 社会交换理论:雇佣关系是一种交换关系,不同类型的交换关系对雇员的工作绩效和态度带来不同的影响。
Miller 和 Shamsie ^[36]	<ul style="list-style-type: none"> ● 资源的作用受环境影响吗? 	 <p>财务资源：独立的财务资源、系统的财务资源</p> <p>知识资源：独立的知识资源、系统的知识资源</p> <p>稳定/动荡的环境</p> <p>财务绩效</p>	<p>历史资料：</p> <ul style="list-style-type: none"> ● 样本：7 家好莱坞大型电影公司。 ● 1936 年到 1950 年,好莱坞电影业处于稳定繁荣的时期,公司发展财务资源,例如与影星签订大量长期合同,在全美各地投资影院; ● 1951 年到 1965 年期间,好莱坞电影业受电视业冲击,大幅震荡.公司发展知识资源,例如研发新技术,增加对综合项目的投资,并减少与影星的长期合同,全面撤资各地影院。 	<p>基于资源基础理论,建立了“资源的权变理论”。</p> <ul style="list-style-type: none"> ● 资源基础理论:企业竞争优势的来源于特殊的异质资源,如财务资源和知识资源。 ● 财务资源受法律保护,使竞争对手无法效仿,在稳定环境中不会过时,在可预测环境中能够产生更优越的财务业绩。 ● 知识资源受知识壁垒保护,竞争对手不知道如何模仿企业的流程或技能,能够帮助公司适应动荡的环境,在动荡环境中能够产生更优越的财务业绩。

3.1 与现实对话以提供基于现实的证据

如果所研究的现象及其关系在客观世界中确实存在,那么就有信心相信构建的理论为真。其一,研究者可以通过看报纸、听新闻、阅读行业或企业报告,获得现实的证据,增强对所构建理论的理解。比如,Tsui等^[35]对员工企业间的雇佣关系的研究,起源于他们观察到1988年到1990年间的美国企业界的一些大的改变。报刊上经常出现一些报道,很多原来在保护员工方面做得很好的企业,为了让企业能够持续下去,不得不开始裁员。他们观察到当时企业与员工间的雇佣关系发生了变化,有些变化不可思议:企业不给员工提供很多保障,但同时还要求员工关心企业,更加努力地做好工作。作者成立讨论小组,每周花两、三小时讨论问题。既看报纸,了解实际情况,又看文献,寻求逻辑解释。经过半年多的持续讨论,他们建立了一个可以接受的员工组织间的雇佣关系模型。

其二,走到现实中观察、访谈,做一些质性研究,也是研究者常用的方法。例如,在Detert和Edmondson^[30]研究隐性建言行为(implicit voice)时,研究者先在一家公司进行问卷调研,发现超过50%的参与者报告说“工作中提出建议、说出想法,常常感到不舒服”。通过深入访谈,发现了员工不愿意建言的5条主要原因。进一步,研究者对185位来自不同组织的高管人员进行问卷调研,发现这5项不愿建言的原因在组织中普遍存在。基于此,研究者开发了一个隐性建言理论。再如, Little等^[28]研究职场怀孕女性如何维持自己的职业形象,提出一个研究模型,即职场女性感知到怀孕带来合法性和价值两方面的威胁,从而积极地采取形象维持策略(如保持做事节奏等)和去类别化策略(如隐瞒怀孕等),来管理职业形象。作者通过对35位职场怀孕女性的访谈,证实了理论所描述的现象和机制的存在,据此发展了“职业形象维持理论”。进一步运用587位怀孕女性的两轮调研数据,分析结果验证了作者所发展的理论。

其三,查阅历史资料,形成故事脉络,也是研究者擅于使用的方法。比如,Miller和Shamsie^[36]研究企业资源与环境的互动对企业财务绩效的影响时,选取7家好莱坞大型电影公司作为研究样

本。通过查阅历史资料,研究者发现,1936年到1950年间,是好莱坞电影公司的“黄金年代”,观众对电影需求旺盛、偏好稳定。这段时期内,这7家电影公司大力发展自身的财务资源。这些稳定的财务资源使得电影公司能够在可预测环境中提前规划各类战略,为电影公司带来了可观收益。而在1951年到1965年期间,电视业的兴起给好莱坞电影业带来了巨大冲击,观影需求减少,偏好差异变大。这段时期内,这7家公司开始增大对知识资源的投入。知识资源的增多使电影公司能够灵活应对外界环境带来的不确定性,使公司财务绩效保持稳定。基于对历史资料的梳理,研究者相信“资源的权变理论”的真实性。

另如,Sherer和Lee^[33]关于制度变革动因的研究中,选取美国律师事务所作为研究样本。该行业长期采取的“非升即走”人才选拔标准使得美国法律界人才短缺情况严重。研究者查阅1970年代的商业杂志,如《美国律师》,发现早期已经有一些律师合伙人对如何改善人才短缺问题进行了探讨,并与资源依赖理论对组织如何解决资源短缺问题相一致。随后,《纽约时报》等商业杂志对美国一家著名的精英法律事务所Davis Polk实施的“高级律师轨道”人才选拔办法进行了大幅报道。新办法逐渐被其它精英事务所效仿,并逐步推广到全美。整合资源依赖理论和合法性逻辑,作者建立了组织变革的内生驱动理论。

再如,Barnett和King^[31]以美国化工行业为对象研究行业为何进行自我监管。1984年3月,美国联合碳化物公司位于印度博帕尔的一家子公司发生化学物质泄露事件,死伤惨重。作者翻阅了大量关于此事件的报道,发现行业内多家公司的都发表过类似看法,“这件事不仅仅是联合碳化物公司一家事情,而是整个行业的事情”,“博帕尔事件很可能导致美国化工企业集体消亡”。媒体报道也表明,博帕尔事件直接促使了美国化学理事会的责任关怀计划的产生。这些历史记录支持了研究者关于现代行业共享非物质资源的观点,表明保护这些非物质共享资源是产生行业自我监管的重要原因之一。

3.2 与理论对话以增强基于理论的逻辑

本文所列举的9篇AMJ年度最佳论文,每篇

都非常强调基于理论的逻辑来构建自己的理论。比如,Chua^[29]基于内隐信念理论,建立起周边文化冲突与跨文化创造力之间的关系。内隐信念是人们在日常生活中用来解释和评估社会环境的意义系统。人们对于来自不同文化的观念和价值观是否相容持有内隐信念。关于文化是否相容的信念既有长期性的一面,也有可变的一面,这些信念也可以被环境刺激暂时改变。因此,作者认为当人们在其直接的社会环境中(例如在他们的社会网络中)目睹跨文化不和谐时,他们认为不同文化的思想和价值观天生不相容的信念就会被激活或增强,从而削弱了整合跨文化想法的能力,破坏了多元文化的创造力。

另如,Agarwal等^[32]基于知识基础观,建立了衍生公司知识继承理论。公司存在两种至关重要的知识:技术知识和市场知识。如果母公司技术知识和市场知识不同步增长,那么拥有这些知识的员工就有创办衍生企业以实现市场价值的动机,从而提高了衍生企业产生的可能性。如果母公司技术知识和市场知识同步增长,那么母公司能够进入许多细分市场,衍生公司存活率小,从而降低衍生企业产生的可能性。衍生公司能够继承母公司的知识遗产,所以衍生公司的存活率高于其他类型的市场新进入者。借助1977年—1997年间磁盘驱动行业的数据,作者验证了所建立的理论。

再如,Seibert等^[34]整合弱连接、结构洞和社会资源三个理论,构建了“职业成功的社会资本理论”,以解释社会资本与个人职业成功之间的关系。弱连接理论认为,弱连接往往是紧密联系的社会小团体之间的桥梁,因此,如果个体在工作场所中拥有更多的弱连接,个体将与其他部门拥有更多的接触,从而获得更多的信息、更多的资源渠道。结构洞理论认为,如果个体在工作场所的网络中拥有更多的结构洞,个体将有更多的机会与上级接触,从而获得更多的信息资源和职业庇护。社会资本理论关注网络中资源的本质。个体在网络中获得的信息、资源和庇护将有利于个体的薪酬、晋升和职业满意度。通过问卷调查,作者获得448位来自不同组织的不同职业的员工数据,验证了所构建的理论。

4 研究结论与意义

4.1 研究结论:混淆统计检验与统计推断之 P 值误用

科学研究是“问题—理论—证据”三位一体的过程,包括推理、检验和推断三项活动。由问题到理论再到假设(H_1),是推理活动;由 H_1 建立 H_0 ,设定 H_0 为真的情况下检验 H_0 与抽样数据匹配概率,是检验活动;基于检验结果来推断 H_0 为假从而接受 H_1 为真,进而接受理论为真,是推断活动。推断的第一步是统计推断。统计推断和统计检验是两个不同的活动。统计检验犯错误的概率叫假阳性概率,即 $\Pr(H_0 \text{ 检验被拒绝} | H_0 \text{ 为真}) = \Pr(P < \alpha | H_0 \text{ 为真}) = \alpha$ 。统计推断犯错误的概率叫假阳性之反概率,即 $\Pr(H_0 \text{ 为真} | H_0 \text{ 检验被拒绝})$ 。假阳性之反概率越小,统计推断的可靠性越高。

“假阳性概率”和“假阳性之反概率”不一样,也不相等。它俩之间的数值关系受假阳性概率(α)、假阴性概率(β)和先验概率(prior probability)的影响,特别是受先验概率影响很大。先验概率越高,利用统计检验的结果(即假阳性概率)进行统计推断所犯错误概率就越小。如果先验概率很小,即使统计检验结果的假阳性概率很小(小于0.05或0.01),则利用该结果进行推断所犯错误概率也很大。

因此, P 值之争的本质是对基于 P 值的NHST方法的误用,把统计检验和统计推断混淆,直接用统计检验结果来进行统计推断。这导致错误的推断。科学研究中,要掌握先验概率,以把握基于统计检验的结果来进行统计推断的可靠性。

本文提出掌握先验概率的双重对话途径:1)与现实对话以提供基于现实的证据;2)与理论对话以增强基于理论的逻辑。并通过9篇*Academy of Management Journal*年度最佳论文具体做法的总结(见表2),指导研究者如何进行。

4.2 掌握先验概率对克服 P 值误用的科学哲学意义

掌握先验概率,对统计检验结果进行修正,再进行统计推断。这具有科学哲学意义。其一,揭示

了科学研究是先验与经验相统一的本质^[37,38]。每一项研究活动不是孤零零的存在,而是与其所在的自然、社会及其已有知识联系在一起。自然、社会及其已有知识提供了研究活动所探究的理論的先验概率,脱离此,统计检验的经验活动没有意义。每项研究活动又作为先验存在进入科学研究的长河之中。

其二,加深了对不充分决定性(underdetermination)的理解。不充分决定性讲的是理论与证据间的关系^[39]。理论代表实在(reality) 经验证据代表事实(factuality) 由于事实是有限的和局部的,所以实在和事实之间不是镜像关系。这导致不能借助任何事实推断出理论正确性。所以,从经验证据到理论的推断过程中,必须考察经验证据存在的情境,获得理论在此情境下发生效力的先验概率。任何研究所接受的理论,若要推广到其他情境,都需要重新审视。

其三,把科学研究看作是通过建立理论来解决问题的活动(见图1),理论是否被接受要看其解题效力^[40]。问题是时空范畴的存在,即使存在超越时空的问题,但对问题考察的经验活动又是时空范畴的存在。所以,科学研究所建立的和所接受的理论,也是时空范畴的存在。建立和接受什么样的理论,主要取决于其在情境中解决问题的效力。在A情境中有效的理论,如果由于在B情境中先验概率很低而无效,那么B情境也不能接受该理论。

4.3 通过双重对话掌握先验概率的实践启示

提示科学研究者和评审者,需要重视以下问题。

第一,研究的是真问题吗?对现象和情境的深刻了解,以及对理论的逻辑及其隐含假设的透彻理解,是抓住真问题的根本方法^[21]。科学界把“用科学的方法解决错误的问题”叫做“第三类错误”。Von Glinow 和 Teagarden^[41]指出,当前中国管理研究中“第三类错误现象突出,其根源在于中国学者对国外理论的边界条件缺乏了解”。与情境和理论的双重对话,可能是避免犯第三类错误的不二之法。

第二,构建理论了吗?通过构建理论来解决问题,是科学工作者有别于管理咨询、管理实践等

从业者的本质特征。发展新理论来解决真问题,是科学工作者的应该所为和尽力所为。无论是发展新理论还是运用旧理论,如果解决的是假问题,这都是不可取的。

第三,有现实证据吗?可以参考上文所列举的AMJ年度最佳论文的做法,比如新闻报道、行业或企业报告,观察、访谈等质性研究,历史资料和故事等,形式多样地提供现实证据。也可以通过元分析来展示所建立理论发挥效力的先验概率。需要谨慎的是,元分析是针对一个具体问题的系统性回顾,其样本情境来源是多样的,而一项研究的统计检验P值是在一定情境中获得的,而需要获得的是该情境下的先验概率,用此先验概率来调整统计检验结果进行统计推断。如果现实证据表明先验概率在0.5或以上,那么可以用统计检验结果进行统计推断;如果低于0.5,那么必须谨慎使用统计检验结果进行统计推断。在报告统计检验结果的时候,不应立即进行统计推断,需要报告先验概率是否在0.5或以上,然后报告统计推断结果。

第四,逻辑坚实吗?Colquitt 和 Zapata-Phelan^[42]给出了假设提出的五种方式,这五种方式反映了“作为实证研究中假设提出的依据,现有理论被应用的程度”。这五种方式有层次之分,从低层次到高层次分别是“用归纳法或逻辑描述法形成预测”、“引用已有发现建立预测”、“基于现有概念说理推导出预测”、“用现有模型和图示形成预测”和“基于现有理论产生预测”。前两种与本文强调的“基于现实的证据”一致,后三者与本文强调的“基于理论的逻辑”相一致。逻辑越坚实,所探究的理論的先验概率越高。

5 结束语

本文认为,P值之争的本质是,对基于P值的零假设检验(NHST)的误解和错用,用统计检验的结果直接进行统计推断,导致似是而非的推断错误。如果理論的先验概率很低,即使统计检验结果显著(如 $P < 0.05$ 或 0.01),利用该结果进行推断所犯错误概率也很高。因此,在根据统计检验结果进行统计推断时,必须考察先验概率。本文提出

掌握先验概率的双重对话途径. 科学工作者应该与情境充分对话, 通过各种方式, 比如阅读报纸、走到现实中访谈、查阅历史资料等, 获得尽可能多的现实知识, 理解理论发生的情境. 同时, 科学工

作者应该与知识体系充分对话, 充分掌握“知识的已知”, 探寻解释现象发生的理论. 双重对话帮助研究者掌握先验概率, 从而提高了基于统计检验结果进行统计推断的可靠性.

参 考 文 献:

- [1] Wasserstein R L, Lazar N A. The ASA statement on p-values: Context, process, and purpose [J]. *The American Statistician*, 2016, 70(2): 129–133.
- [2] Rozeboom W W. The fallacy of the null-hypothesis significance test [J]. *Psychological Bulletin*, 1960, 57(5): 416–428.
- [3] Goodman S. A Dirty Dozen: Twelve P-Value Misconceptions [C]. *Seminars in Hematology*, WB Saunders, 2008, 45(3): 135–140.
- [4] Greenland S, Senn S J, Rothman K J, et al. Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations [J]. *European Journal of Epidemiology*, 2016, 31(4): 337–350.
- [5] Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance [J]. *Nature*, 2019, 567(7748): 305–307.
- [6] Wasserstein R L, Schirm A L, Lazar N A. Moving to a world beyond “ $p < 0.05$ ” [J]. *The American Statistician*, 2019, 73(sup1): 1–19.
- [7] Wagenmakers E J, Wetzels R, Borsboom D, et al. Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem [J]. *Journal of Personality and Social Psychology*, 2011, 100(3): 426–432.
- [8] 胡传鹏, 孔祥祯, Eric-Jan Wagenmakers, 等. 贝叶斯因子及其在 JASP 中的实现 [J]. *心理科学进展*, 2018, 26(6): 951–965.
Hu Chuanpeng, Kong Xiangzhen, Eric-Jan Wagenmakers, et al. The Bayes factor and its implementation in JASP: A practical primer [J]. *Advances in Psychological Science*, 2018, 26(6): 951–965. (in Chinese)
- [9] Valentine J C, Aloe A M, Lau T S. Life after NHST: How to describe your data without “p-ing” everywhere [J]. *Basic and Applied Social Psychology*, 2015, 37(5): 260–273.
- [10] 钟建军, Dienese Z, 陈中永. 心理研究引入贝叶斯统计推断的必要性, 应用思路与领域 [J]. *心理科学*, 2017, 40(6): 1477–1482.
Zhong Jianjun, Dienese Z, Chen Zhongyong. Bayesian inference be introduced to the psychological research: Its necessity, ideas and field [J]. *Journal of Psychological Science*, 2017, 40(6): 1477–1482. (in Chinese)
- [11] Open Science Collaboration. Estimating the reproducibility of psychological science [J]. *Science*, 2015, 349(6251): 1–8.
- [12] Johnson V E. Revised standards for statistical evidence [J]. *Proceedings of the National Academy of Sciences*, 2013, 110(48): 19313–19317.
- [13] Benjamin D J, Berger J O, Johannesson M, et al. Redefine statistical significance [J]. *Nature Human Behaviour*, 2018, 2(1): 6–10.
- [14] McShane B B, Gal D, Gelman A, et al. Abandon statistical significance [J]. *The American Statistician*, 2019, 73(sup1): 235–245.
- [15] Simmons J P, Nelson L D, Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant [J]. *Psychological Science*, 2011, 22(11): 1359–1366.
- [16] Lakens D, Scheel A M, Isager P M. Equivalence testing for psychological research: A tutorial [J]. *Advances in Methods and Practices in Psychological Science*, 2018, 1(2): 259–269.
- [17] Benjamin D J, Berger J O. Three recommendations for improving the use of p-values [J]. *The American Statistician*, 2019, 73(sup1): 186–191.

- [18] Dewey J. How We Think [M]. Boston: D. C. Heath & Co. , 1910.
- [19] Kerlinger F N , Lee H B. Foundations of Behavioral Research (4th ed) [M]. Fort Worth: Harcourt College Publishers , 2000.
- [20] Smith K G , Hitt M A. Great Minds in Management: The Process of Theory Development [M]. New York: Oxford University Press , 2005.
- [21] 盛昭瀚. 问题导向: 管理理论发展的推动力 [J]. 管理科学学报 , 2019 , 22(5) : 1 – 11.
Sheng Zhaohan. Problem orientation: Driving force of management theory development [J]. Journal of Management Sciences in China , 2019 , 22(5) : 1 – 11. (in Chinese)
- [22] Whetten D A. What constitutes a theoretical contribution? [J]. Academy of Management Review , 1989 , 14(4) : 490 – 495.
- [23] Fisher R A. The Design of Experiments (9th ed.) [M]. Edinburgh: Oliver and Boyd , 1935.
- [24] Neyman J , Pearson E S. On the problem of the most efficient tests of statistical hypotheses [J]. Philosophical Transactions of the Royal Society of London , Series A , Containing Papers of a Mathematical or Physical Character , 1933 , (231) : 289 – 337.
- [25] Schwab D P. Research Methods for Organizational Studies [M]. Mahwah: Lawrence Erlbaum , 2005.
- [26] 丹尼尔·卡尼曼. 思考, 快与慢 [M]. 北京: 中信出版社 , 2012.
Kahneman D. Thinking , Fast and Slow [M]. Beijing: CITIC Press , 2012. (in Chinese)
- [27] Bem D J. Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect [J]. Journal of Personality and Social Psychology , 2011 , 100(3) : 407 – 425.
- [28] Little L M , Major V S , Hinojosa A S , et al. Professional image maintenance: How women navigate pregnancy in the workplace [J]. Academy of Management Journal , 2015 , 58(1) : 8 – 37.
- [29] Chua R Y J. The costs of ambient cultural disharmony: Indirect intercultural conflicts in social environment undermine creativity [J]. Academy of Management Journal , 2013 , 56(6) : 1545 – 1577.
- [30] Detert J R , Edmondson A C. Implicit voice theories: Taken-for-granted rules of self-censorship at work [J]. Academy of Management Journal , 2011 , 54(3) : 461 – 488.
- [31] Barnett M L , King A A. Good fences make good neighbors: A longitudinal analysis of an industry self-regulatory institution [J]. Academy of Management Journal , 2008 , 51(6) : 1150 – 1170.
- [32] Agarwal R , Echambadi R , Franco A M , et al. Knowledge transfer through inheritance: Spin-out generation , development , and survival [J]. Academy of Management Journal , 2004 , 47(4) : 501 – 522.
- [33] Sherer P D , Lee K. Institutional change in large law firms: A resource dependency and institutional perspective [J]. Academy of Management Journal , 2002 , 45(1) : 102 – 119.
- [34] Seibert S E , Kraimer M L , Liden R C. A social capital theory of career success [J]. Academy of Management Journal , 2001 , 44(2) : 219 – 237.
- [35] Tsui A S , Pearce J L , Porter L W , et al. Alternative approaches to the employee-organization relationship: Does investment in employees pay off? [J]. Academy of Management Journal , 1997 , 40(5) : 1089 – 1121.
- [36] Miller D , Shamsie J. The resource-based view of the firm in two environments: The Hollywood film studios from 1936 to 1965 [J]. Academy of Management Journal , 1996 , 39(3) : 519 – 543.
- [37] 孟 强. 研究作为科学的本质——先验视角与经验视角 [J]. 南京社会科学 , 2020 , (6) : 16 – 22.
Meng Qiang. Research as essence of science: Transcendental and empirical perspectives [J]. Nanjing Journal of Social Science , 2020 , (6) : 16 – 22. (in Chinese)
- [38] 许晓东, 陈 玲. 近代“经验-先验”观念的转变与科学革命 [J]. 自然辩证法研究 , 2021 , 37(8) : 98 – 103.
Xu Xiaodong , Chen Ling. The change of the concept of “Experience-prior” and the scientific revolution in modern times [J]. Studies in Dialectics of Nature , 2021 , 37(8) : 98 – 103. (in Chinese)
- [39] 徐淑英, 李绪红, 贾良定, 等. 负责任的管理研究——哲学与实践 [M]. 北京: 北京大学出版社 , 2018.
Xu Shuying , Li Xuhong , Jia Liangding , et al. Responsible Research in Management: Philosophy and Practice [M]. Beijing: Peking University Press , 2018. (in Chinese)

- [40] 曾 点. 从“科学研究纲领”到“科学研究传统”——劳丹的非本质主义哲学观[J]. 自然辩证法研究, 2017, (3): 22–26.
Zeng Dian. From “Research Program” to “Research Tradition”: Laudan’s philosophical non-essentialism[J]. Studies in Dialectics of Nature, 2017, (3): 22–26. (in Chinese)
- [41] Von Glinow M A, Teagarden M B. The future of Chinese management research: Rigour and relevance redux[J]. Management and Organization Review, 2009, 5(1): 75–89.
- [42] Colquitt J A, Zapata-Phelan C P. Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal[J]. Academy of Management Journal, 2007, 50(6): 1281–1303.

The P -value debate and implications for management research: The significance of prior probability

JIA Liang-ding¹, LIN Tse-min², WANG Tao³, BU Meng-meng⁴, HE Gang⁵

1. School of Business, Nanjing University, Nanjing 210093;
2. Department of Government, Department of Statistics and Data Sciences, University of Texas at Austin, Austin 78712, USA;
3. Emlyon Business School, Department of Strategy & Organization, Ecully 69130, France;
4. School of Business Administration, Nanjing University of Finance and Economics, Nanjing 210023, China;
5. Guanghua School of Management, Peking University, Beijing 100871, China

Abstract: This paper argues that the P -value debate is grounded in the confusion between statistical significance testing and statistical inferences. Using results of statistical significance testing to make statistical inferences can lead to the validation of many specious theories. Errors in statistical significance testing refer to the probability of false positives, and errors in statistical inferences lie in the inverse probability of false positives. The relationship between the two probabilities is highly affected by prior probability. A higher prior probability decreases the probability of errors when using results of statistical significance testing for statistical inferences. By contrast, a lower prior probability increases the probability of statistical inference errors even if the probability of false positives is low. To improve the reliability of statistical inferences, and taking management research as an example, this paper proposes to establish a better prior probability through a dual dialogue with practical and theoretical worlds. The significance of this paper lies in revealing the unity of prior knowledge and empirical evidence in management research and the insufficiently deterministic nature of management research.

Key words: P -value; null hypothesis significance testing (NHST); inverse probability of false positives; prior probability; management research