改进统计生态学的四项原则

Gordana Popovic[1,6], Tanya J. Mason[2], Tiago A. Marques[3,4], Joanne Potts[5], Szymon M. Drobniak[6,7], Rocío Joo[8], Res Altwegg[9], Carolyn C. I. Burns[10], Michael A. McCarthy[11], Alison Johnston[12], Shinichi Nakagawa[6], Louise McMillan[13], Kadambari Devarajan[14,15], Patrick l. Taggart[16], Alison C. Wunderlich[17], Magdalena M. Mair[18,19], Juan Andrés Martínez-Lanfranco[20], Malgorzata Lagisz[6], Patrice P. Pottier[6]

1. Stats Central，Mark Wainwright 分析中心，悉尼新南威尔士大学，澳大利亚 2. 生态系统科学中心，生物、地球与环境科学学院，悉尼新南威尔士大学，澳大利亚 3. 生态与环境建模研究中心，天文台，圣安德鲁斯大学，圣安德鲁斯，苏格兰 4. 统计与应用中心，动物生物学系，里斯本大学科学学院，葡萄牙 5. The Analytical Edge 统计咨询公司，邮政信箱 47 号，布莱克曼斯湾，塔斯马尼亚，澳大利亚 6. 进化与生态学研究中心，生物、地球与环境科学学院，新南威尔士大学，悉尼，澳大利亚 7. 环境科学研究所，雅盖隆大学，克拉科夫，波兰 8. 全球渔业观察，华盛顿特区 20036，美国 9. 生态、环境与保护统计中心，统计科学系，开普敦大学，7701 朗德博斯，南非 10. 悉尼，澳大利亚 11. 生态系统与森林科学学院，墨尔本大学，帕克维尔，维多利亚州，澳大利亚 12. 生态与环境建模研究中心，数学与统计学，圣安德鲁斯大学，圣安德鲁斯，英国 13. 数学与统计学院，惠灵顿维多利亚大学，惠灵顿，新西兰 14. 生物体与进化生物学研究生项目，马萨诸塞大学阿默斯特分校，阿默斯特，马萨诸塞州，美国 15. 自然资源科学系，罗德岛大学，金斯敦，罗得岛州，美国 16. 脊椎动物害兽研究组，新南威尔士州初级产业部，昆比恩，新南威尔士州，澳大利亚 17. 生物科学研究所，圣保罗州立大学，海岸校区，圣维森特，圣保罗，巴西 18. 统计生态毒理学，拜罗伊特大学，拜罗伊特，德国 19. 理论生态学，雷根斯堡大学，雷根斯堡，德国 20. 生物科学系，阿尔伯塔大学。埃德蒙顿。阿尔伯塔省，加拿大

摘要

近年来，统计方法的误用日益受到关注，尤其令人担忧的是，不良实验设计、挑选性报告和报告不充分等做法十分普遍。这些问题大多并非有意为之，在生态学中也不比其他科学学科更常见；只要给予恰当指导，其中许多问题都很容易得到纠正。

本文源于2020年国际统计生态学会议上的一次讨论，阐述生态学家如何遵循四项指导原则开展研究，以实现更具影响力的统计研究实践：1. 明确聚焦的研究问题，然后规划抽样和分析以回答该问题；2. 建立能够体现数据分布和依赖性的模型；3. 强调效应量，以生态学相关性取代统计显著性；4. 充分详细地报告方法和研究发现，使研究有效且可重复。

这些原则按大致重要性顺序排列，为实验设计和报告提供了一个框架，有助于防范不可靠的实践。从明确界定研究问题入手，研究人员便能设计高效的研究来回答问题，并避免导致假阳性和可重复性不佳的不良研究实践。恰当且正确的统计模型能够得出可靠结论；良好的报告实践以及对生态学相关性的关注，则能使结果更具影响力且可重复。

本文以一项近期关于干扰对高地沼泽影响的研究为例，阐释选择和运用有效统计实践的依据，并为希望改进统计方法运用的生态学家提供实用指导。

关键词

统计学原则、结果已知后提出假设、p值操纵、生态学相关性、统计显著性、可重复性、研究浪费、假设

引言

报告研究发现时，生态学家和其他科学家一样，希望研究结果能反映所研究系统中真实发生的情况，并传达其生态学相关性和证据强度。为了使研究结果经得起检验，研究人员需要遵循良好的研究实践。未能遵循良好实践，导致许多领域的研究发现可重复性较低（开放科学合作组织，2015年；Begley 和 Ellis，2012年；Camerer 等，2018年）。生态学中也普遍存在不良研究实践（Anderson 等，2000年；Fidler 等，2006年；Fraser 等，2018年），采用这些做法可能会扭曲研究发现、浪费资源、未能充分报告生态群落中发生的情况，最终还可能向其他研究人员、政策制定者和公众歪曲研究结果。

糟糕的研究实践大多源于对统计方法和研究设计原则的误解或曲解。最常见且影响最深远的此类做法包括：

- 结果已知后提出假设（HARKing；Kerr，1998），据报告，51%的生态学家和进化生物学家采用过这种做法（Fraser 等，2018年）。
- 不报告不显著结果；这是一种挑选性报告，64%的生态学家承认自己至少做过一次（Fraser 等，2018年）。
- 基于事先已知为假的零假设进行假设检验。Anderson 等，2000年发现，他们评估的《生态学》文章中，绝大多数（95%）所含的零假设很可能事先就已知为假。
- 将不显著结果误解为“没有效应”或“没有关系”的证据；在生态学领域报告不显著结果时，约有63%的情况会出现这种误解（Fidler 等，2006年）。
- 方法和分析细节说明不足。近 80% 的生态学论文未能提供足够的细节，无法实现计算上的可重复性（其中大多数论文，即 73%，未附上分析代码；Culina等人，2020）。

面对这份成绩单，生态学家可能会辩称，他们的研究实践相对无害，或是不可避免的。然而，对假设检验作用的误解，会使任何此类检验及其结果失去意义。此外，众所周知，挑选性报告和结果已知后提出假设会（大幅）抬高假阳性率（即总体中实际并无效应，却得出“显著”结果；例如参见 Forstmeier 等，2017），导致许多错误结果发表。不幸的是，我们并不知道生态学中假阳性的真实比例，因为重复研究十分罕见；迄今发表的所有研究中，真正的重复研究仅占 0.023% （ Kelly，2019 ），尽管生态学家和进化生物学家普遍认为重复非常或比较重要（ 97% 的受访者；Fraser等人，2018）。不过，我们确实知道，生态学研究中有相当高的比例（70%）支持其原始假设（Fanelli，2010），这表明其中许多结果可能是假阳性。另一方面，研究浪费也不容忽视。Purgar 等人（2021）发现，生态学领域 82% 至 89% 的研究似乎都因低质量研究、发表偏倚以及研究设计、分析和报告不佳而造成了本可避免的浪费。

我们以四项原则为依据阐述统计学最佳实践，希望借此引导生态学研究者，尤其是刚入门的研究者，通过避免这些常见陷阱，为自己的结论提供尽可能有力的证据。我们确定的四项原则是：

1. 首先，明确一个聚焦的研究问题，然后规划抽样和分析，以回答该问题。2. 建立一个能够考虑数据分布和依赖性的模型。

3. 强调效应量，以生态学相关性取代统计显著性。4. 充分详细地报告方法和研究发现，使研究有效且可重复。

这些原则大致按照重要性和影响力排序，后面的原则要发挥作用，需要以先前原则奠定的基础为依托。例如，如果模型没有考虑依赖性，报告效应量也无法改善研究。在收集任何数据之前明确聚焦的研究问题，可以杜绝结果已知后提出假设；若再配合注册，还能进一步避免挑选性报告。制定抽样计划有助于更好地回答研究问题，从而获得更有意义、影响更深远的研究结果；采用更好的抽样设计，也能提高发现真实效应的可能性（即统计功效）。正确地对数据建模同样能够提高统计功效并控制假阳性率。强调效应量能够将重点放在生态学相关性上，而这正是生态学研究最有意义的结果。全面报告方法和研究发现，有助于他人成功复现你的研究，使你的研究发现更具分量，并推动该领域向前发展。

.

这四项原则源自2020年国际统计生态学会议期间的讨论。构思本文的讨论小组成员来自不同领域，都是目前在野外工作的生态统计学家。我们注意到，尽管已有许多介绍统计学良好实践的优秀文献，其中也包括生态学领域的文献，但这些文献往往侧重于开展特定分析的流程（例如：Zuur 等人，2010年；Steel 等人，2013年；Zuur 与 Ieno，2016）或解决特定问题（Nakagawa 与 Parker，2015），而不是提出一小组易于理解、适用于所有分析的原则。通过阐明这些原则，我们希望赋能生态学家开展更加稳健且有意义的研究，并通过帮助建立共同的研究方法论。来促进生态统计学领域的合作。

在全文中，我们将通过一个生态学案例来阐释这些原则，探讨干扰如何影响高地沼泽（Mason 等人，2022）。众所周知，地下采矿会扰乱地表水流和地下水流，进而可能影响附近的沼泽群落。研究人员希望探讨不同的水分可利用性如何影响沼泽植物群落，以及这种影响是否会因火灾扰动而改变。本研究从多个沼泽中采集了中宇宙样本，随后将其随机分配到温室中的水分处理和火处理组。本研究所说的中宇宙，是指一柱土壤和植物：将 PVC 管（直径 150 毫米、深 250 毫米）敲入地面至与地面齐平后，再用沟渠铲挖出。随后将这些样本放入温室的水槽中，并调节水槽水位，以模拟不同的地下水可利用性。20 个月后，通过依次移除生物量（修剪）、施加热量和烟雾，在每种水分处理下的一半中宇宙中模拟火灾事件（见图 1）。

在每个章节中，我们都会介绍适用于各项原则的实用 R 软件包（R Core Team，2020）。我们选择 R 语言，是因为它是开放的免费软件，也是生态学领域使用最广泛的软件（Lai 等人，2019）。

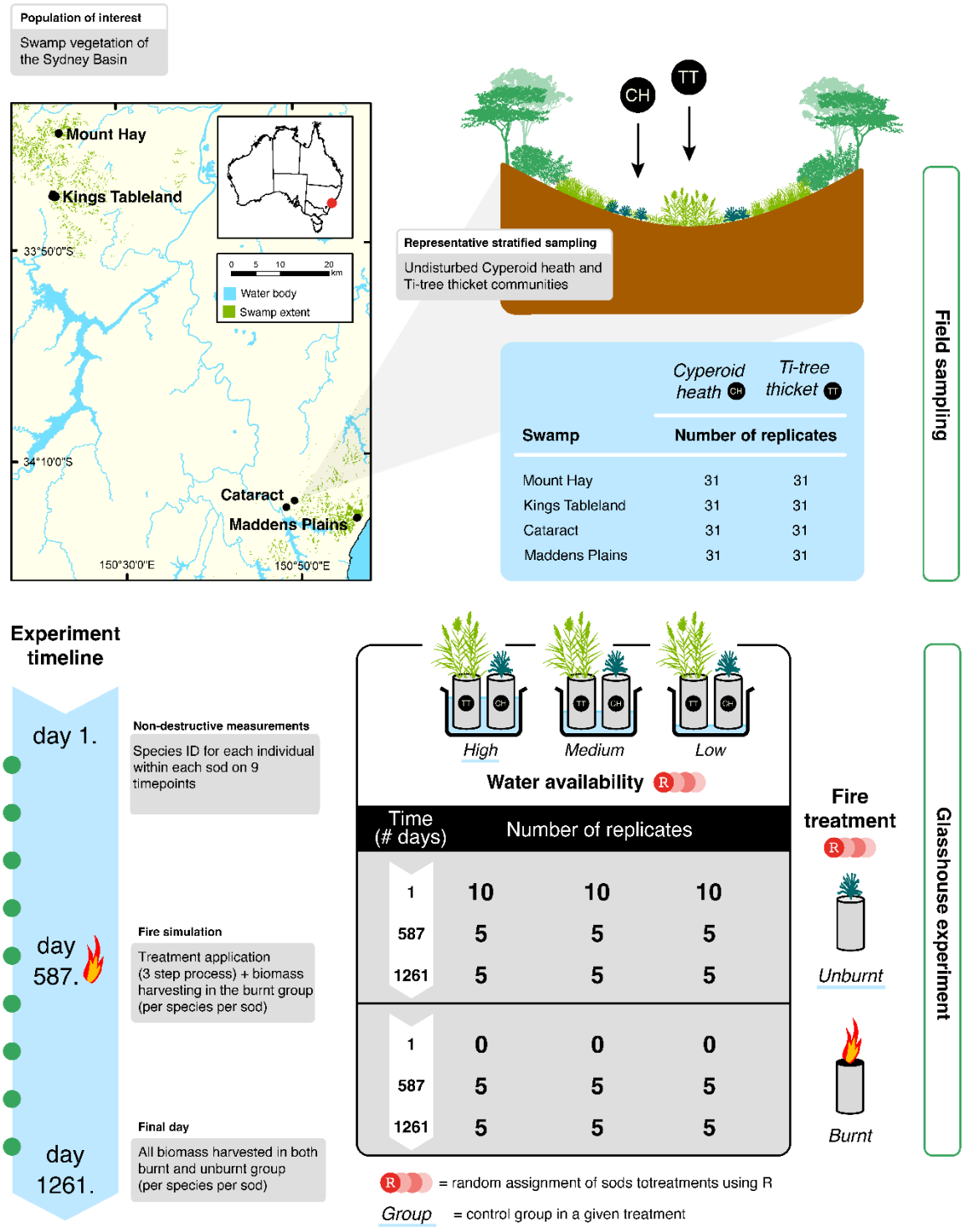

图 *1* – 沼泽研究的实验设计

原则 1：首先明确一个聚焦的研究问题，然后规划抽样和分析，以回答该问题

1A – 明确你的研究问题

提出一个好的研究问题是研究过程中最重要的环节。构思不周的问题不仅会降低研究发现的实用性，还可能在整个研究过程中引发方法学问题。好的研究问题应当符合可行、有趣、新颖、合乎伦理且相关（FINER标准；Hulley，2007），而且每个研究问题都应对应一份具体的统计分析计划，理想情况下还应对其进行预注册（见第 1C 节）。在沼泽示例中，研究人员认为温室实验既可行又合乎伦理；文献综述证实其具有新颖性；而且，鉴于沼泽下方的长壁采煤规划许可是湿地管理中一个现实且争议颇大的议题，这项研究既有趣又相关。

PICO框架（Haynes，2006年）常用于构建健康与医学研究中的研究问题，但也很容易调整后用于生态学研究。对于每个研究问题，都应明确以下内容：

- **总体**——感兴趣的总体或目标总体，通常根据物种、空间和时间来界定。
- **干预**——随机实验中施加于研究对象的处理，以及观察性研究中关注的解释变量。
- **比较**——明确计划采用什么作为参照或对照组，以便与处理组进行比较。
- **结局**——指计划测量哪些结局，以考察干预的有效性或关注的解释变量所产生的影响，通常称为响应。在生态学中，这些指标可以包括丰度、丰富度、位置、速度和多样性等。

尝试预测研究中会发生什么通常很有帮助，因为这有助于明确研究问题。沼泽研究人员认为，可利用水分较少的中宇宙，其生物量和丰富度会低于水分可利用性较高的中宇宙；而火会加剧低水分的影响。本例中感兴趣的总体是澳大利亚悉尼盆地的高地沼泽植物群落。干预组由水位和火的不同组合构成：水分可利用性分为低、中、高三个水平；火的水平分为烧过和未烧过。对照组为高水位中宇宙。结局包括生物量、丰富度，以及每个物种的存在/缺失（物种组成）。

在收集数据*之前* 明确界定研究问题（或探索先前收集的数据），是确保实验设计稳健的最重要步骤。明确界定研究问题的好处显而易见，但为何在开展研究后不应细化或修订假设，可能就不那么明显了。在收集与原始研究问题相关的数据后再提出假设，会扭曲研究结果，造成一种假象，让人误以为支持*事后*假设的证据非常有力（Forstmeier 等，2017）。这是因为，收集数据后对假设所做的任何改动，都可能反映样本中的模式，而这些模式未必能反映总体中的模式，从而增加研究结果显示假阳性的可能性。

然而，禁止*事后*提出假设，并不妨碍研究人员开展探索性分析，以为未来的研究问题和假设提供参考；事实上，这可能是

确定后续研究方向的最佳方式之一。作为生态学家，我们常常会对非常复杂的数据进行概括，以便用简单的模型来分析。例如，在沼泽研究中，你可以将多物种数据汇总为一个丰富度指标，以回答首要研究问题。然而，这种复杂性对于理解研究物种及其群落的生态学往往非常重要；花时间探索原始数据，常常会带来新的发现和研究方向。数据探索可以采取绘图、制表、模型拟合和假设检验（包括*p*值）等形式。每当你在没有预先设定研究问题的情况下探索数据时，都应将其视为（并报告为）探索性分析或假设生成（Forstmeier 等，2017）。

1B – 使数据收集与研究目标相匹配

明确研究问题后，下一步是确定回答这些问题需要哪些数据。数据收集可能包括观察、抽样和实验操纵。操纵实验是证明因果关系的唯一方法，只要可行，就应始终开展。然而，操纵实验在生态学研究中确实存在缺点：其范围有限（受时间限制，以及处理的数量和类型限制），而且可能不符合伦理（例如，造成具有重大负面影响的环境扰动）。操纵实验的情境也可能不够真实，例如，沼泽研究受到限制，因为在温室中无法让新物种进入中宇宙。最好将操纵实验与自然实验和纵向监测结合起来（Driscoll 等，2010）。

抽样

具有代表性的样本能够反映目标总体的模式和特征。在沼泽示例中，目标总体由悉尼盆地未受干扰的高地沼泽群落中最湿润的亚群落（茶树灌丛和莎草状石南荒地）里的物种组成。

在野外收集新数据时，代表性样本是确保无偏推断的最佳方式。生态学复杂的时空特性可能使代表性抽样设计面临挑战，生态学家过去曾接受代表性较差甚至随意数据收集的方式（Smith 等，2017）。我们提倡生态学家使用代表性样本，以获得最可靠且无偏的结果。有用的工具包括基础 R 中的 `sample()` 函数，以及 `sp` 包中的 `spsample()` 函数（Pebesma 和 Bivand，2005）。

如果由于野外工作中意外出现的挑战或存在偏差的既有数据集，未能实现代表性抽样，则可通过在建模中使用权重和偏移量（用于校正已知偏差，例如抽样强度）、协变量（用于校正与测量变量相关的不平衡，例如地点可达性）以及相关结构（见原则 2A），来处理非代表性抽样。或者，您可以将推断限定于您的样本所代表的总体（Williams 和 Brown，2019）。与收集新的野外数据相比，在处理既有数据集时，重要的是考虑其中任何固有偏差。使用他人收集的数据时，最佳做法是联系并与收集这些数据的研究人员合作，因为他们最了解自己的数据。通过谨慎选择用于分析的数据，可以近似实现代表性抽样设计（Johnston 等，2021）。

实验、相关性与因果关系

研究人员通常最关注因果关系，以回答某个响应是否以及如何受到一组给定协变量的促进、导致或诱发。发现这种因果关系可能是揭示潜在生态机制的第一步。

唯一能够确凿证明因果关系的方法是开展实验，在实验中研究人员操控环境并观察其后果。实验必须具备：

1. 对照（了解不进行操控时会发生什么） 2. 重复（在各处理组内进行；将观察到的差异归因于操控，而非随机变异） 3. 随机化（将采样单位分配到处理组；避免偏差以及与未测量／未控制变量的混杂）

沼泽示例展示了如何在复杂环境中实现这一点。从多个沼泽采集了中宇宙，随后随机分配到不同处理组，使研究人员能够展示水和火处理与植物群落之间的因果关系。研究设置了若干处理组合，其中高水位、未燃烧的中宇宙作为对照组。通过抽样并将多个中宇宙分配到每种处理中，实现了重复。采样单位（中宇宙）需要随机分配到各处理组（例如低、中、高水位）。这种随机化可以借助广泛可用的软件轻松实现，例如`randomizr`软件包（Coppock，2019），沼泽示例中采用的就是这种方法；也可以简单地从帽子里抽取号码。

许多设计严谨的实验还会采用盲法：负责收集测量数据或分析数据的研究人员不知道处理分配情况。盲法可以消除无意识的观察者偏倚，但遗憾的是，在生态学领域并未广泛采用；符合条件的生态学研究中，实际采用盲法的仅约占 13%（Kardish et al., 2015年）。

对于生态学家，尤其是野外生态学家而言，开展对照实验往往并不可行。次优选择是在回归中控制混杂变量，或采用其他“因果”分析方法（例如 Larsen et al., 2019）。然而，只有在附带限定条件的情况下，才能将回归和因果分析的结果作因果解释，因为这些方法依赖于关于数据的非常强的假设，而这些假设通常无法检验。任何因果方法的使用都应附有详细讨论，说明所有假设及其合理性。

1C – 制定分析计划并考虑注册

在收集或探索现有数据之前，制定一份可靠的分析计划至关重要。尽早规划分析方法，可以显著降低最终分析的复杂性，并使结果更加清晰。有效的数据分析计划可确保研究能够回应研究问题，并与采样方法同步制定（关于如何最好地建立能够考虑数据特征的模型，见原则 2）。

分析计划必须全面列出拟开展的模型和检验，并逐一说明模型类型（例如，采用高斯分布的混合效应模型）、结局/响应变量（例如，丰富度的对数）、关注的主要效应（例如，水位），以及需要控制的变量（例如，沼泽和植被群落的固定效应、中宇宙的随机效应）。通常，我们无法预先确定哪种模型最适合数据；在这种情况下，应说明如何选择合适的模型。在沼泽示例中，研究人员首先假设丰富度服从高斯分布，然后检查残差图；如果高斯分布不合适，就改用对数 + 1变换。进行多重检验时，应采用适当的方法控制多重检验（例如参见 Pike，2011），以控制假阳性。

为获得最稳健的实验设计，我们建议注册研究（我们遵循 Rice & Moher，2019年的做法，倾向于使用“注册”一词而非“预注册”）。注册会在数据收集之前，有时也会在数据分析之前，将对研究的详细描述存档之前（Nosek 等人，2018年；Parker 等人，2019；Rice & Moher，2019）。这类描述包括研究目标（原则 1A：研究问题、假设和预测）、研究设计（原则 1B；数据收集过程）以及数据分析计划（原则 2A：拟合的统计模型）。注册可通过公共注册平台加上时间戳，例如开放科学框架或*As Predicted*；如有需要，也可以设置禁运期。 注册报告与注册类似，因为二者都会在数据收集前确定研究假设和分析计划，但它是一种独特的程序（Chambers，2013），会在数据收集前对引言和方法进行同行评审。目前，生态学期刊正在考虑采用注册报告，例如*BMC Ecology and Evolution*和*Ecology and Evolution* 以及 *Conservation Biology*；一些专业性较低的期刊也在考虑采用，例如《*PLOS One*》、*BMC Biology, Scientific Reports* 以及 *Nature Communications*。

注册可以对研究质量产生重大影响。举一个来自其他领域的例子，最近一项研究表明，生物医学和心理学研究中的注册报告仅支持了约 40% 的原始假设（Allen & Mehler，2019），而传统文献中的比例为 80%–95%。尽管目前生态学领域尚无与之对应的研究，但我们预计，注册和注册报告会使生态学中的正面与负面研究发现比例更加合理。

原则 2：建立一个能够解释数据分布和依赖性的模型

2A – 模型依赖性

误差独立性是所有统计方法的一项关键假设。误差是指对依赖性和协变量影响进行建模后，数据中尚未建模的部分。独立性假设可以通过收集独立数据（例如，采用简单随机样本）来满足，也可以通过适当建模（解释）数据中的任何依赖性来满足。

依赖性通常源于抽样设计。例如，考虑沼泽示例中那种分层（或嵌套）抽样设计，其中每个实验单元（沼泽）都进行了子抽样。依赖性之所以产生，是因为同一沼泽中的中宇宙彼此之间比不同沼泽中的中宇宙更相似。这种子抽样（通常称为“伪重复”）是依赖性的常见来源（Hurlbert，1984年）；时间依赖性也是如此，例如反复测量某个中宇宙中的物种丰度。此类重复测量（Gurevitch & Chester，1986年）同样会导致同一中宇宙中的观测值彼此之间比不同中宇宙中的观测值更相似。空间依赖性也十分常见，指的是空间上较近的地点比相距较远的地点更相似。依赖性也可能源于系统的复杂特性。例如，在多物种调查中，物种丰度之间的依赖性可能来自对未观测环境梯度的共同响应、系统发育或物种间的相互作用。

依赖性本身并不是问题；如果理解得当，重复观测可以极具价值，但在分析中必须正确处理（Steel 等，2013）。如果未能正确处理依赖性，就会导致错误的推断（通常是假阳性）。大多数统计软件都能轻松地对依赖性进行建模（详细指南见 Warton，2022）。简而言之，广义线性（混合）模型（McCulloch 与 Neuhaus，2006）可通过 `lme4` 实现

（Bates等，2015，第4页）软件包可以通过为聚类变量（~ （1 |中宇宙））加入随机截距，对聚类导致的依赖性进行建模。对于多变量依赖，可使用 `gllvm`（Niku等，2020）和 `Hmsc`（Tikhonov等，2020）软件包实现的层级模型，也可以使用广义估计方程（例如 `mvabund`；Wang等，2012）。在沼泽示例中，存在多个层次的依赖性。同一片沼泽中的中宇宙彼此依赖，对每个中宇宙的重复测量彼此依赖，而且测量涵盖多个物种，这些物种之间存在复杂的相互依赖关系。沼泽研究人员使用 `glmmTMB`（Brooks等，2017）拟合了层级模型；除了关注的效应（植被、水和火处理）之外，他们还纳入了沼泽的固定效应、中宇宙的随机效应，以及物种间的降秩相关结构（见图2）。

2B – 检查假设

我们已在第2A节讨论过对依赖性进行建模的重要性。此外，统计模型假设响应来自某个特定的概率分布，该分布具有若干关键属性，包括响应类型（连续型、二元型、计数）及其均值—方差关系。例如，连续型数据（如体重）适用高斯（正态）分布，该分布假设方差恒定；存在/缺失数据适用二项分布，其响应为二元型，且均值—方差关系呈二次关系；计数数据（如丰度）适用泊松分布或负二项分布，它们分别假设方差等于或大于均值。我们不应试图让数据适应模型，而应努力构建能够体现数据特征的模型。如果不满足模型假设，模型就无法体现数据的特征，研究人员应尽力寻找拟合效果更好的模型。违反模型假设会使参数估计值产生偏差，例如混淆位置效应和离散效应（Warton等，2012）；还会增加I类错误（假阳性研究发现；Queen等，2002），从而导致分析得出错误结论。如果找不到更合适的模型（可能是因为不存在这样的模型，或者正确的模型尚未在代码中实现），就应进行敏感性分析。敏感性分析就是尝试采用具有不同假设的备选模型，以评估分析结论是否能经受这些变化（即结果是否一致）。

每个拟合出的模型都必须经过检查并进行调整，之后才能得出结论。通常通过目视检查残差图来检验假设。残差与预测值图可以显示是否偏离线性或假定的均值—方差关系。绘制残差随时间或空间变化的图，可以分别显示是否存在时间自相关或空间自相关。正态分位数图可以诊断是否偏离正态性（如果使用正态分布）。不同模型的假设及其对违反假设的敏感程度各不相同。例如，在线性回归中，偏离正态性没有方差恒定那么关键；违反方差恒定可能会提高第一类错误率（Glass等，1972）。由于生态学家拟合的大多数分布都是离散的（例如泊松分布、负二项分布、二项分布），残差有许多种选择。在`statmod`（Dunn和Smyth，1996）和`DHARMa`（Hartig，2020）软件包中实现的分位数残差，或许最适合生态学家用于检验假设。

有时，为了对数据进行建模，我们必须作出无法检验的假设。常见的两种情形是缺失数据，此时需要对缺失机制的类型作出假设（Nakagawa 与 Freckleton，2008）；以及根据观察性数据进行因果推断。在这种情况下，敏感性分析有助于评估结果对这些假设有多敏感。

原则 3：强调效应量，以生态学相关性取代统计显著性

3A – 强调效应量，以生态学相关性取代统计显著性

生态学家希望找到一些工具，帮助他们谈论研究发现的生态学“显著性”，或者用一个更恰当且不易引起混淆的术语，评估研究发现的“生态学相关性”（Martínez-Abraín，2008年）。由于这很难做到，生态学家往往退而求其次，只说结果“具有统计显著性”。虽然这很诱人，但统计显著性不足以用于报告生态建模的结果，因为它无法表明效应量（Nakagawa & Cuthill，2007年）。

假设总体中的效应量微乎其微——例如，两种水处理之间的平均生物量相差 0.01%——只要充分增加样本量，就总能达到统计显著性。但这并不会让 0.01% 的差异变得更具生态学相关性。效应量（此处为平均差）才是我们关注的量，在分析和报告中必须优先考虑；同时还应讨论所测效应量在生态学上有多大相关性，而这必须以专家知识为依据。

注重效应量的估计，要求报告并解读置信区间（或在采用其他统计方法时使用的等效指标，例如贝叶斯分析中的可信区间）。这些区间的宽度表示估计值所伴随的不确定性。相比显著性检验，如果注重估计，就能更好地把握统计分析与解读中的细微差异。

在沼泽示例中，研究考察了多种效应。研究人员关注不同水处理对生物量、丰富度和组成的影响，因此分析重点是对照组（水位较高）与另外两个水位（低水位和中水位）之间的差异。对于火灾处理，研究人员关注其与水处理之间的交互作用，即在燃烧过和未燃烧的中宇宙之间，水位对生物量、物种丰富度和丰度的影响有何差异。例如，作者假设，在未燃烧的处理中，生物量会随着水位降低而减少，而在燃烧过的处理中，这种效应会更加明显。

3B – 弱化 *p* 值

p 值的定义是：在总体中不存在效应或差异（零假设）的前提下，获得与样本中观察到的结果相同或更极端结果的概率。这一定义冗长复杂，也难怪它常常遭到误解。本质上，这意味着*p*值可以表明数据与指定统计模型的不相容程度。*p* 值并非零假设为真的概率，也不能表示效应大小，或数据仅由随机偶然性产生的概率；此外还有其他误解（Wasserstein & Lazar, 2016）。

有人呼吁彻底弃用*p*值和假设检验，原因既包括它们常被错误解读，也包括前文提到的问题，例如挑选性报告，以及检验那些先验已知为假的假设。更普遍的建议是，统计学家认为我们应停止将*p*值二分（分为显著／不显著；McShane 等人，2019）；这种做法不仅混淆了统计显著性与生态学显著性，也是造成这些不良

做法（例如*p*值操纵）的原因之一。统计学家还建议降低*p*值的重要性。当前的最佳实践是，将*p*值作为效应量和置信区间等更重要指标的补充进行报告，这也会降低*p*值被误解的可能性。正如 Wasserstein等人（2019，第2页）所建议的，解读结果时最好"接受不确定性。保持审慎、开放和谦逊"。

在沼泽示例中对生物量随时间的变化进行建模时，研究人员发现"有证据（t_{341} = 3.093，P = 0.006）表明，未燃烧的低水位与高水位中宇宙之间的生物量差异随时间推移而增大；实验第587天至第1261天期间，高水位与低水位中宇宙之间的生物量差异增加了一倍以上（变化= 2.183；95% 置信区间：1.205–3.955）"。置信区间表明，生物量差异可能小至增加20%（1.205），也可能大至增加至原来的四倍（3.955）。前者或许可视为生态学相关性较小，而增加一倍或更多则可能被视为效应较大。研究人员还发现，"没有证据表明高水位与中水位中宇宙之间的生物量变化存在差异（t_{341} = 0.320，P = 0.945；变化= 1.085；95% 置信区间：0.597–1.971）"。较宽的95%置信区间（0.597–1.971）表明，高水位与中水位中宇宙之间的差异可能意味着生物量减半（0.5），也可能意味着增加一倍（2.0）。较大的*p*值和较宽的置信区间，可能是样本量过小或方差过大造成的，也可能是高水位与中水位中宇宙之间确实不存在差异。研究人员无法确定这些因素中哪一个导致了较大的*p*值，因此无法断定高水位与中水位之间不存在差异。

原则4：详细报告方法和研究发现，使研究具有说服力且可重复。

4A – 让研究发现易于重复

报告研究时，可重复性和重复性都是重要的考量因素。如果对同一数据集执行相同的分析步骤，始终得到相同的答案，那么结果就是可重复的。如果对不同的数据集执行相同的分析，得到的答案在定性上相似，那么结果就是可重复的（The Turing Way 社区，2023）。如前所述，生态学领域的重复研究比例非常低（Kelly，2019），而且大多数已发表的研究没有充分详述研究设计、数据收集和分析，以至于根本无法开展重复研究（Culina等人，2020）。协作网络或许是提高生态学领域重复研究数量的有效途径，例如近期的营养网络、美国长期生态研究网络和大叶藻实验网络。全面报告除了有益于科学研究之外，成功重复你的研究还能大大增强研究发现的分量，而这正是你应当追求的目标。你应提供他人重复你的研究时可能需要的所有信息，帮助他们完成重复研究。

如果你以前从未在发表的论文中附上数据和代码，这件事可能会让人感到无从下手，但它并不是非此即彼的选择；朝着可重复性迈出的每一步都很有价值。我们的建议遵循同事 D. Falster（个人通信）的观点：从小处着手，逐步实现完全可重复；首先，编写代码，确保过一段时间后你自己仍能复现结果；其次，确保合作者能够复现这些结果；最后，只需再多做一点工作，就能在发表的论文中提供完全可重复的代码。以下是一些较好的实践，难度大致按由低到高排列：

- 始终以"非专有的机器可读格式提供完整数据（即不要提供扫描的纸质页面或 PDF 文件）；
- 记录研究日志，注明研究设计或分析的任何变更及其原因，并在方法和补充材料中报告这些内容；
- 确保代码可执行——也就是说，脚本文件或 R Markdown 文档应能从头到尾顺利运行，不出现中断、异常或错误；
- 为论文附上完整代码，以便复现研究结果（如果采用非代码方法，则提供详细的工作流程，以追溯分析步骤）；
- 尽可能避免使用付费软件，并推广开源且可免费使用的计算工具；
- 避免使用短暂且不稳定的托管平台来存放代码和数据（例如个人网站、院系网络档案）。改用免费、公开且维护良好的存储库（例如 Figshare、Dryad、Zenodo）；
- 展示代码时，使用版本控制系统（如 Git）以及整合代码与注释的丰富文档（例如使用 R Markdown 和 R 软件包，如 *knitr*）；
- 声明分析中使用的程序包及其版本（例如，借助`renv`程序包）。如果所用的软件或程序包可能很快发生变化，导致难以重复使用，可以考虑将代码和数据打包成一个独立的软件包（例如使用 Docker）。

4B – 将模型输出可视化以传达结果

正如我们在原则 3 中所讨论的，结果应侧重于效应量和置信区间，以促进对生态学相关性的关注。实现这一点最有效的方法之一，是绘制模型输出。模型输出指模型产生的效应估计值和置信区间。虽然有些软件包内置了相应函数，但`emmeans`软件包（Lenth，2021）可以绘制几乎所有常用 R 软件包的模型输出。另一种方法是使用大多数 R 软件包中的`predict()`函数，计算带有不确定性的预测，然后手动绘图。绘制带有不确定性（即置信区间）的模型输出十分重要；在`emmeans`中使用`Cis = TRUE`，或在许多软件包的`predict()`函数中使用`se.fit = TRUE` ，即可轻松实现这一点。沼泽研究中的图表使用`emmeans`绘制（代码见下文），然后用`ggplot`进行了修改（Wickham，2016）；这里重现了丰富度随时间变化的模型估计值和置信区间（图2）。

```
library(lme4); library(emmeans)
rich_mod <- lmer(丰富度 ~ 沼泽 + 植被 + 时间 * 水 +
                                    水 * 火 + 时间 * 火 +
                                    (1 | 中宇宙), 数据 = 丰富度)
emmip(rich_mod, ~ 水 ~ 时间 | 火, Cis = TRUE)
```

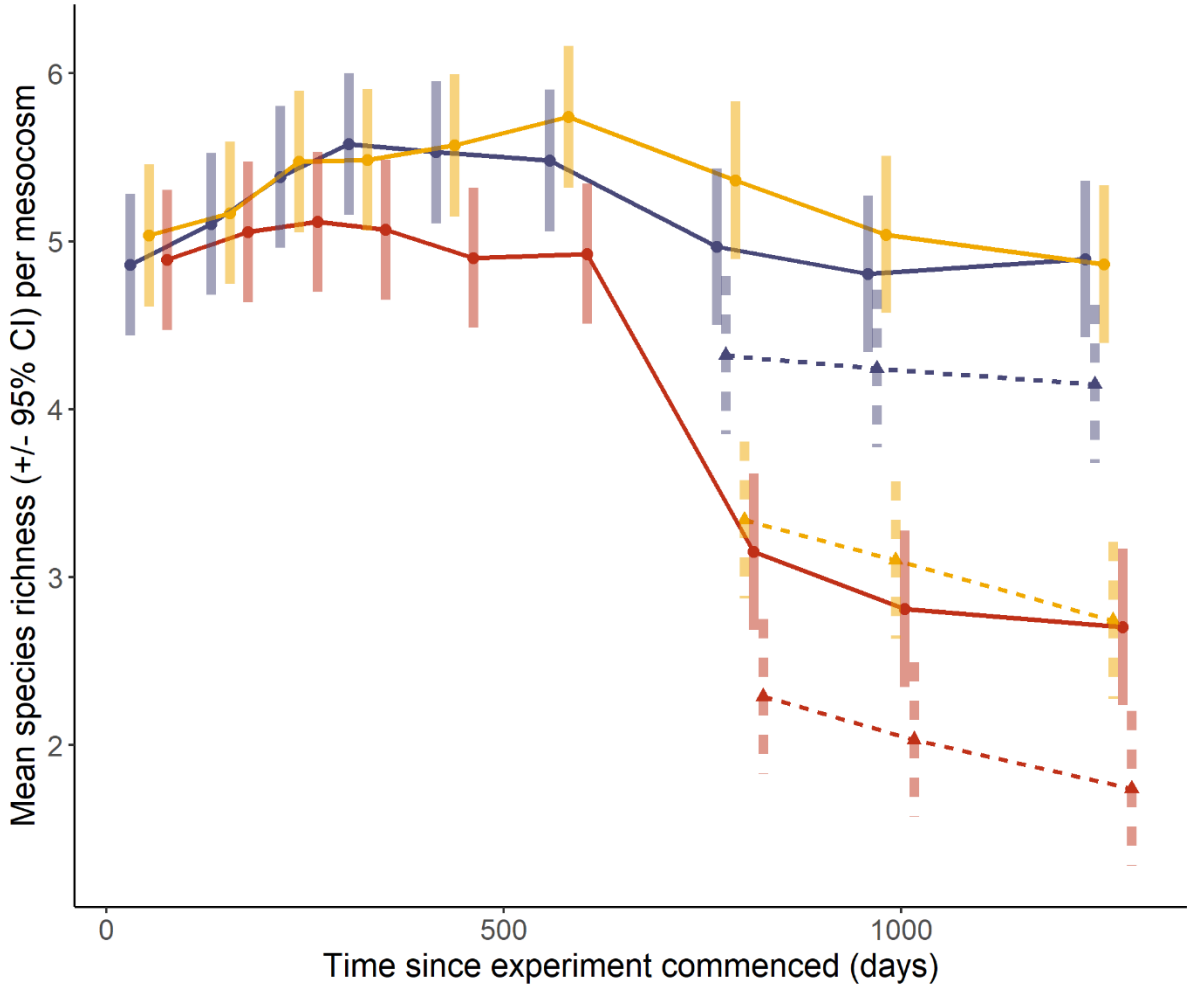

图2：处理组中宇宙中物种丰富度随时间（开始后的天数）变化的估计边际均值（*± 95%* 置信区间）。处理水平包括高（）、中（）和低（）水分可利用性，以及未燃烧（）和燃烧（）火灾处理水平。

讨论

本文提出了旨在改进生态学研究统计方法的指南，供个人研究人员和研究团队使用。尽管超出了本文的讨论范围，但毫无疑问，在学科的系统层面上，还可以采取更多措施，推动更完善的实验设计和报告实践得到广泛采用；作者也了解到，一些期刊已开始在评审流程中纳入更严格的方法学要求。我们承认，当研究人员全身心投入各自的研究领域时，要持续关注统计实践的复杂性可能颇具挑战，

我们希望，对统计方法提供清晰的指导并开展批判性运用，有助于提升统计能力和熟练度，从而使生态学研究群体受益。

作者贡献

GP 提出研究构想，并与 CB 共同牵头撰写和综合。TJM 提供了沼泽示例。TAM、JP、SB、RJ、RA、MAM、AJ、SN 和 LM 牵头撰写各部分内容（作者顺序为随机排列）。KD、PIT、ACT、ACW、MMM、JAM、ML 和 PPP 参与了多个部分的撰写（作者顺序为随机排列）。所有作者都对稿件草案提出了重要意见，并最终批准发表。

资助

TJM 感谢新南威尔士州政府通过其环境信托基金（2018/SSC/0049）和拯救我们的物种计划提供的协助。TAM 感谢 CEAUL 提供的部分支持（由葡萄牙科学与技术基金会 FCT 通过项目 UIDB/00006/2020 资助）。RA 获南非国家研究基金会支持（资助编号：114696）。ACW 获圣保罗研究基金会支持（资助编号：#17/16650–5）。MMM 获克里斯蒂安娜·纽斯莱因-沃尔哈德基金会支持。JAM 获国家研究与创新局（乌拉圭国家研究与创新局）以及计算生物多样性科学与服务计划（Bios2-Canada）支持。PP 获新南威尔士大学 Scientia 博士奖学金支持。

致谢

作者谨向所有参与 vISEC2020 讨论组及后续讨论的人士表示感谢。感谢 Glenda Wardle、Annemieke Drost、Nilanjan Chatterjee、Pedro Nicolau、Chloe Bracis、Teresa Neeman、Javier Seoane、Sarah Marley、Noa Rigoudy、Brenton Annan、Rebecca Groenewegen、Theresa O'Brien、Michelle Marraffini、Julie Vercelloni、Andrea Havron、Hayden Schilling、Amanda Hart、Christine Stawitz、Fabiana Ferracina、Andrew Edwards、Mick Wu、Gesa von Hirschheydt、Rick Camp、Alison Ketz、Julie Vercelloni、Sarah Saunders 和 Sarah Hasnain 提供的头脑风暴贡献。此外，作者感谢 Frederic Gosselin 指出了用于检验不可信的零假设的重要参考文献，并感谢 Daniel Falster 提供了有关可重复性的建议。

参考文献

Allen, C., & Mehler, D. M. A. (2019). 职业生涯早期及以后阶段的开放科学挑战、益处与建议。*PLOS Biology*, *17*(5), e3000246. https://doi.org/10.1371/journal.pbio.3000246 Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). 零假设检验：问题、普遍性及一种替代方案。*The Journal of Wildlife Management*,*64*(4), 912. https://doi.org/10.2307/3803199 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). 使用 lme4 拟合线性混合效应模型。*Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01 Begley, C. G., & Ellis, L. M. (2012). 提高临床前癌症研究的标准。*Nature*,*483*(7391), 531–533. https://doi.org/10.1038/483531a Brooks, M. E., Kristensen, K., Benthem, K. J. van, Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). GlmmTMB 在速度与

零膨胀广义线性混合建模软件包之间的灵活性方面实现了平衡。*TheR Journal*, *9*(2), 378–400. Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., ... Wu, H. (2018). 评估2010年至2015年间发表于《自然》和《科学》的社会科学实验的可重复性。*Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z Chambers, C. D. (2013). 注册报告：Cortex 推出的一项全新出版计划。*Cortex*,*49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016 Coppock, A. (2019). *randomizr*：用于常见随机分配和抽样方式的易用工具。https://CRAN.R-project.org/package=randomizr Culina, A., van den Berg, I., Evans, S., & Sánchez-Tójar, A. (2020). 生态学中代码的可用性很低：呼吁立即采取行动。*PLOS Biology*, *18*(7), e3000763. https://doi.org/10.1371/journal.pbio.3000763 Driscoll, D. A., Lindenmayer, D. B., Bennett, A. F., Bode, M., Bradstock, R. A., Cary, G. J., Clarke, M. F., Dexter, N., Fensham, R., Friend, G., Gill, M., James, S., Kay, G., Keith, D. A., MacGregor, C., Russell-Smith, J., Salt, D., Watson, J. E. M., Williams, R. J., & York, A. (2010). 生物多样性保护中的火灾管理：关键研究问题及我们回答这些问题的能力。*Biological Conservation*, *143*(9), 1928–1939. https://doi.org/10.1016/j.biocon.2010.05.026 Fanelli, D. (2010). “正面”研究结果沿科学等级体系逐级增加。*PLOS One*,*5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068 Fidler, F., Burgman, M. A., Cumming, G., Buttrose, R., & Thomason, N. (2006). 对零假设显著性检验的批评对保护生物学统计报告实践的影响。*Conservation Biology*, *20*(5), 1539–1544. https://doi.org/10.1111/j.1523-1739.2006.00525.x Forstmeier, W., Wagenmakers, E., & Parker, T. H. (2017). 识别并避免可能的假阳性研究发现——实用指南。*Biological Reviews*, *92*(4), 1941–1968. https://doi.org/10.1111/brv.12315 Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). 生态学与进化领域中的可疑研究做法。*PLOS One*, *13*(7), e0200303. https://doi.org/10.1371/journal.pone.0200303 Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972年). 未能满足方差与协方差固定效应分析所依据假设的后果。*Review of Educational Research*, *42*(3), 52. Gurevitch, J., & Chester, S. T. (1986年). 重复测量实验的分析。《生态学》,*67*(1), 251–255. https://doi.org/10.2307/1938525 Hartig, F. (2020). *DHARMa*：层级（多层/混合）回归模型的残差诊断。https://CRAN.R-project.org/package=DHARMa Haynes, R. B. (2006). 构建研究问题。*Journal of Clinical Epidemiology*, *59*(9), 881–886. https://doi.org/10.1016/j.jclinepi.2006.06.006 Hulley, S. B. (2007). 临床研究设计。Lippincott Williams & Wilkins. Hurlbert, S. H. (1984年). 伪重复与生态学野外实验设计。*Ecological Monographs*, *54*(2), 187–211. Johnston, A., Hochachka, W. M., Strimas-Mackey, M. E., Ruiz Gutierrez, V., Robinson, O. J., Miller, E. T., Auer, T., Kelling, S. T., & Fink, D. (2021). 提高

社区科学数据的价值：利用 eBird 数据估算物种分布的实例。《多样性与分布》，*27*(7), 1265–1277. https://doi.org/10.1111/ddi.13271 Kardish, M. R., Mueller, U. G., Amador-Vargas, S., Dietrich, E. I., Ma, R., Barrett, B., & Fang, C.-C. (2015). 生态学、进化与行为学中对未盲法观察的盲目信任。《生态学与进化前沿》，*3*. https://doi.org/10.3389/fevo.2015.00051 Kelly, C. D. (2019). 生态学与进化研究的重复率与成功率。*PeerJ*, *7*, e7654. https://doi.org/10.7717/peerj.7654 Kerr, N. L. (1998). 结果已知后提出假设（HARKing）：结果已知后提出假设。《人格与社会心理学评论》，*2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4Lai, J., Lortie, C. J., Muenchen, R. A., Yang, J., & Ma, K. (2019). 评估 R 语言在生态学中的流行程度。*Ecosphere*, *10*(1). https://doi.org/10.1002/ecs2.2567 Larsen, A. E., Meng, K., & Kendall, B. E. (2019). 利用观察性数据对对照-影响生态学研究进行因果分析。生态学与进化方法, *10*(7), 924–934. https://doi.org/10.1111/2041-210X.13190 Lenth, R. V. (2021). *emmeans*：估计边际均值，又称最小二乘均值。https://CRAN.R-project.org/package=emmeans Martínez-Abraín, A. (2008). 统计显著性与生物学意义：呼吁对生态学研究结果作出更谨慎的解释。《生态学学报》，*34*(1), 9–11. https://doi.org/10.1016/j.actao.2008.02.004 Mason, T., Popovic, G., McGillycuddy, M., & Keith, D. (2022). 水文变化对易发生火灾的湿地植被的影响：实证模拟。《生态学杂志》。https://doi.org/10.1101/2022.06.17.496658 McCulloch, C. E., & Neuhaus, J. M. (2006). 广义线性混合模型。约翰·威立父子公司，纽约。https://doi.org/10.1002/9780470057339.vag009.pub2 McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). 摒弃统计显著性。《美国统计学家》，*73*(sup1), 235–245. https://doi.org/10.1080/00031305.2018.1527253 Nakagawa, S., & Cuthill, I. C. (2007). 效应量、置信区间与统计显著性：生物学家的实用指南。《生物学评论》，*82*(4), 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x Nakagawa, S., & Freckleton, R. P. (2008). 缺失与不作为：忽视缺失数据的危险。《生态学与进化趋势》，*23*(11), 592–596. https://doi.org/10.1016/j.tree.2008.06.014 Nakagawa, S., & Parker, T. H. (2015). 重复生态学与进化研究：可行性、激励机制与成本效益难题。*BMC Biology*, *13*(1), 88. https://doi.org/10.1186/s12915-015-0196-3 Niku, J., Brooks, W., Herliansyah, R., Hui, F. K. C., Taskinen, S., Warton, D. I., & Veen, B. van der. (2020). *gllvm*：广义线性潜变量模型。https://CRAN-R-project.org/package=gllvm Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). 预注册革命。《美国国家科学院院刊》，*115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114 开放科学合作组织。（2015）。估算心理科学的可重复性。《科学》，*349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Parker, T., Fraser, H., & Nakagawa, S. (2019). 通过预注册和注册报告提升保护科学的可靠性。*Conservation Biology*, *33*(4), 747–750. https://doi.org/10.1111/cobi.13342 Pebesma, E., & Bivand, R. S. (2005). 空间数据的类与方法：sp 软件包。*RNews*, *5 (2)*。https://cran.r-project.org/doc/Rnews/ Pike, N. (2011). 在生态学与进化研究的多重比较中使用错误发现率：多重比较中的错误发现率。生态学与进化方法, *2*(3), 278–282. https://doi.org/10.1111/j.2041-210X.2010.00061.x Purgar, M., Klanjscek, T., & Culina, A. (2021). 识别、量化、行动：挖掘生态学研究尚未利用的潜力。*EcoEvoRxiv*。https://doi.org/doi:10.32942/osf.io/xqshu Queen, J. P., Quinn, G. P., & Keough, M. J. (2002). 生物学家的实验设计与数据分析。剑桥大学出版社。R Core Team. (2020). *R*：统计计算语言与环境。R 统计计算基金会。https://www.R-project.org/ Rice, D. B., & Moher, D. (2019). 限制"预注册"一词的使用：一个误用术语。心理科学展望, *14*(6), 1105–1108. https://doi.org/10.1177/1745691619858427 Smith, A. N. H., Anderson, M. J., & Pawley, M. D. M. (2017). 生态学家能否采用更随机的抽样？空间抽样中随意抽样的简单替代方案。*Ecography*,*40*(11), 1251–1255. https://doi.org/10.1111/ecog.02821 Steel, E. A., Kennedy, M. C., Cunningham, P. G., & Stanovick, J. S. (2013). 生态学中的应用统计学：常见陷阱与简单解决方案。*Ecosphere*, *4*(9), art115. https://doi.org/10.1890/ES13-00160.1 The Turing Way 社区。（2023）。*The Turing Way*：可重复、合乎伦理且协作式研究手册。Zenodo。http://doi.org/10.5281/zenodo.3233853 Tikhonov, G., Ovaskainen, O., Oksanen, J., de Jonge, M., Opedal, O., & Dallas, T. (2020).*Hmsc*：物种群落层级模型。R 软件包版本 3.0-9。Wang, Y., Naumann, U., Wright, S., & Warton, D. (2012). *mvabund*：分析多变量丰度数据的统计方法。Warton, D. I. (2022). *Eco-Stats*：生态学中的数据分析：从 *t* 检验到多变量丰度。施普林格国际出版。https://doi.org/10.1007/978-3-030- 88443-7 Warton, D. I., Wright, S. T., & Wang, Y. (2012). 基于距离的多变量分析混淆了位置效应和离散效应。生态学与进化方法, *3*(1), 89–101. Wasserstein, R. L., & Lazar, N. A. (2016). ASA 关于*p* 值的声明。《美国统计学家》, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108 Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). 迈向超越"$p < 0.05$"的世界。《美国统计学家》, *73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913 Wickham, H. (2016). *ggplot2*：优雅的数据分析图形。施普林格纽约。https://ggplot2.tidyverse.org Williams, B. K., & Brown, E. D. (2019). 生态学推断的抽样与分析框架。生态学与进化方法, *10*(11), 1832–1842. https://doi.org/10.1111/2041-210X.13279

Zuur, A. F., & Ieno, E. N. (2016). 回归类分析结果的实施与呈现规范。生态学与进化方法，*7*(6), 636–645. https://doi.org/10.1111/2041-210X.12577 Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010年). 避免常见统计问题的数据探索规范：数据探索。生态学与进化方法，*1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x