

网上摘录：相关关系和因果关系

《新英格兰医学杂志》曾刊登过一篇论文，声称吃巧克力可以提高认知功能。得出这一结论的根据是每个国家的诺贝尔奖得主数量与该国巧克力人均消费量具有强相关性。这靠谱吗？

2012 年，《新英格兰医学杂志》刊登了一篇论文，声称吃巧克力可以提高认知功能。得出这一结论的根据是每个国家的诺贝尔奖得主数量与该国巧克力人均消费量具有强相关性。令人诧异的是这篇论文竟然通过了同行评审，因为很显然作者犯了两个常见错误，这是生物医学文献中研究者进行相关分析时常犯的错误。

相关性描述了两个现象之间的线性关系强度（为了简化说明，这里着重阐述最常用的线性关系——皮尔森相关）：例如，一个变量值的增加可能会跟随着另一个变量值的增加；或者可能是负相关，一个变量的增加会伴随另一个变量的减少。它们的关系是用一个无单位的数值来计算的，那就是相关系数，其数值范围是从-1 到+1，绝对值越接近 1，表示相关性越强。

错误：集体相关=个体相关

《新英格兰医学杂志》那篇论文犯的第一个错误是在群体数据的基础上得出关于个体的结论，陷入了生态学谬误中。这个案例中，作者计算了集合层面（国家）的相关系数，然后错误地把这个数值用于得出关于个体层面的结论（吃巧克力提高认知功能）。实际上，个体层面的准确数据完全是未知的：没有人收集过诺贝尔奖得主吃过多少巧克力的数据，甚至于他们到底是否吃过。事实上，加利福尼亚大学圣迭戈分校的比阿特丽斯·戈隆布（Beatrice A. Golomb）的研究团队检验了这个假设，指出两个变量之间根本不存在相关性。

该论文从未被撤回，至今已被引用 23 次。即使错误的论文被撤回，相关新闻报道还保留在互联网上，还能继续传播错误信息。如果这些对统计分析有所误解的错误结论甚至能出现在《新英格兰医学杂志》这样的著名期刊，那么人们会很好奇，这样的错误出现在生物医学文献中的总体频率有多大？

错误：相关关系=因果关系

第二种曲解相关性分析的错误更常见，即把相关性等同于因果关系。两个因素之间表现出一种关系，可能不是意味着它们之间相互影响，而是意味它们都被同一个隐藏的因素影响。在巧克力消费量和诺贝尔奖得主的例子中，可能是一个国家的富裕程度同时影响了巧克力的消费量和高等教育的可获取性。由此可见，完全无关的现象也可能会可笑地呈现出相关性。相关性可以很肯定地指出一个可能存在的因果关系，但是并不足以证明存在这样的因果关系。

即使科学家很清楚相关性不等于因果关系，但是把两者混为一谈的研究在顶级期刊中还是很普遍。比如，一项 1999 年发表在《自然》杂志上的研究发现，两岁以下儿童的近视和睡眠期间夜间环境光曝光度存在强相关。但是，2000 年发表在《自然》杂志上的另一项研究结果驳斥了上述发现，指出儿童近视的原因是遗传性的，而不是环境因素造成的。这个新的研究发现父母近视与儿童近视之间存在强相关，指出近视的父母为孩子的卧室夜间开灯的可能性更大。在这个例子中，作者基于假想的关联得出了一个结论，而没有检查其他可能的解释。