

实验二 相关研究中的因果判断

相关分析可以提供变量间的相关系数，但一般说来，研究者无法仅由此来确定两个变量间是否存在因果关系，以及因果关系的方向。在一般的相关研究中，轻易地做出一个因素导致另一个因素的结论是十分草率地，因为相关系数能告诉我们的仅是两个变量间的关联程度。

但在另一些相关研究中，研究者也试图通过一些逻辑推论来加强相关研究的解释力，使其具有部分程度上的因果推断能力。此前介绍过的交叉—滞后—组相关法便是这样的一种复相关程序，这一技术依据随时间变化的若干相关系数的大小和方向，来确定是什么因素导致了什么结果。其潜在假设是：既然任何因果效应的发生都需要占用一定的时间，那么如果一个变量引起了另一个变量，则第一个变量与时间上滞后的第二个变量之间的相关应该较大，而它和同期的第二个变量之间的相关则相对较小。本章征文中介绍的埃伦及其同事对儿童对暴力性电视节目的爱好和其攻击性的相关研究，就反映了交叉-滞后程序的这一潜在逻辑。

你可以在一个新的例子中运用类似的逻辑推论，从相关研究中找到因果线索吗？下表是 11 个国家相隔 20 年的香烟消费量与肺癌死亡人数：

表 1 11 个国家中 20 年前的香烟消费量与 20 年后的死亡率

国 家	1930 香烟消费量	1950 死亡人数/百万*
澳 大 利 亚	480	180
加 拿 大	500	150
丹 麦	380	170
芬 兰	1100	350
英 国	1100	460
荷 兰	490	240
冰 岛	230	60
挪 威	250	90
瑞 典	300	110
瑞 士	510	250
美 国	1300	200

*：肺癌患者

（采自多尔，1955）

想一想，这两组数据将会说明什么问题？请根据皮尔逊 r 公式

$$r = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

计算出这两个变量之间准确的相关系数。你得到的相关系数是怎样的，其大小与符号各是什么？根据这个相关系数，你会得出吸烟导致肺癌的结论吗？如果相关系数比较大，比如说+0.95，你是否会更加坚信其间存在着因果关系？如果你认为这些数据并不能证明吸烟会导致肺癌，那么你怎样解释这些结果？请结合前面的例子，对这项研究的结果作出恰当的解释。

（采自坎特威茨等，2001）