



下载歌曲



Search input field

签到 客户端

推广加币 招募版主 升级VIP 升级贵宾

注册|登录

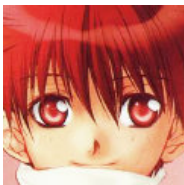
论坛 经济学人 二区 学术道德监督 【热门】统计显著性和经济学家的节操

Stata培训 EViews培训 SPSS培训 《Hadoop大数据分析师》现场&远程	金融建模 CFA/FRM 投行估值 商务图表 CVA注册估值分析师 VBA金融 并购估值	R语言数据挖掘_实战案例分享现场培训 【Stata】高级计量经济学与Stata_陈强主讲	【CDA数据分析师】LEVEL 2培训 spss论文分析不求人 SSC论文视频
SAS培训 Matlab培训 R培训 AMOS培训 【MATLAB基础+金融应用】现场班	统计年鉴汇总 行业分析报告汇总 【SAS培训】金融业的SAS实际运用特训	数据定制服务 数据处理与分析 成为VIP会员 成为贵宾下载更轻松	中国数据分析和数据挖掘俱乐部(CDA)邀请参 人大经济论坛考研专业课辅导班
CDA数据分析师认证 Matlab初中高级培训 《SAS/SPSS数据统计分析师》认证培训	量化投资:思想,策略与R语言基础+实战 【高薪】CDA大数据分析师就业班(包会)	淘宝购物赚论坛币 Python案例分析现场班 【SPSS】进阶SPSS应用与统计分析现场培训	【高薪】CDA数据分析师脱产就业班 Excel图表之道 Excel远程培训

发帖 1 2 下一页 返回列表

提升主题 | 本版置顶 | 关闭主题 | 变更主题颜色 | 抢沙发 | 顶贴 | 显身卡 | 道具中心

楼主: iRolly 【学术治理与讨论】【热门】统计显著性和经济学家的节操 [推广有奖] 343 10



加关注 串个门
加好友 发消息

0 关注 | 27 粉丝

内部工作人员

教授

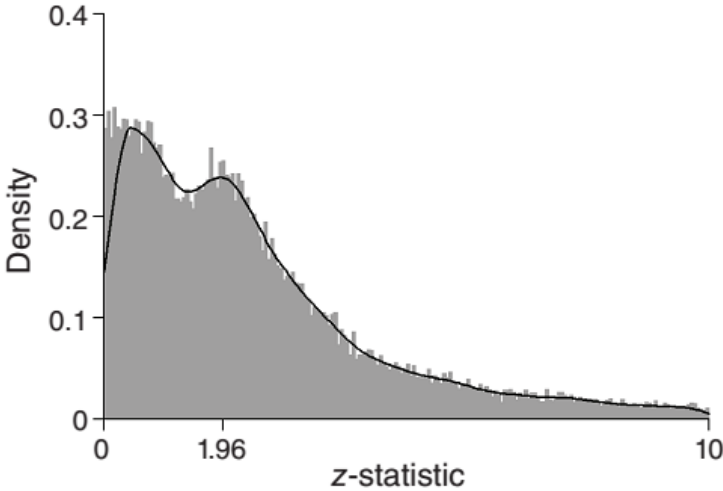
还不是VIP/贵宾 74%

威望 2 级
论坛币 17217 个

iRolly 发表于 2016-3-2 21:34:03 | 只看作者 | 倒序 楼主

本帖最后由 iRolly 于 2016-3-2 21:39 编辑

人大经济论坛经管爱问微信好号“jgasker”好文共享与推荐，实时答疑服务，欢迎关注！



完全免费的高顿CFA题库
PC、手机都可快速做题，APP
也可下载，好东西值得分享
cfa.gaodun.cn

数据分析全体系都在这里
从零学习数据分析
需要怎样的体系？CDA告诉你！
<http://www.cda.cn/>

如果你碰巧有个经济学家朋友，而且热(zuo)心(si)地询问过他最近的研究进展如何，那你多半听到过这样的回答：“唉，做不出来啊。”如果你碰巧又对他们这个行当有所了解，那你就心知肚明，这句回答背后的潜台词是：“老子觉得自己这个idea牛X爆了，结果怎么就他喵的弄不显著呢！？”

伴随着计量技术的飞速发展，主流经济学期刊对于实证结果的要求越来越高，而统计显著性，几乎成为了实证文章得到发表的先决条件。在回归表格里看到标志着显著性的小星星（一般来说，一颗星代表在10%水平上显著，两颗星5%，三颗星1%），差不多是所有实证经济学家梦寐以求的目标。随便翻开一篇发表在主流经济学

期刊上的实证论文，跳到结果部分，最先跃入你眼帘的，就是数不清的星星，一颗，两颗，三颗……

好东西太多了也会惹人厌烦，星星亦是如此。学界一直有人抱怨：我们是不是进入了一个星星“通货膨胀”的时代？一个关于 MIT 的经济学家，达隆·阿西莫格鲁的著名笑话是：他的结果太显著，电脑里的星星都不够用了（见 Acemoglu Facts，一个以黑阿西莫格鲁为主要事业的网站）。

太多的星星真的是一件坏事吗？毕竟统计显著性标志着结果的可信程度，主流期刊偏好可信程度高的研究，似乎也无可厚非。然而，真正令人担心的是，期刊对于星星的重视，可能会激励学者在论文中只提及那些统计显著的结果，甚至会利用一些统计技巧，让原本不显著的结果变得显著。如果真是这样，我们恐怕就不得不为经济学家朋友们的节操感到隐隐的担忧了。

在今年最新一期《美国经济学杂志：应用经济学》（American Economic Journal: Applied Economics）上，四位学者发表了一篇以《星战：实证研究的反击》（Star Wars: The Empirics Strike Back）为题的论文，对上述可能性进行了考察。他们收集了2005到2011年间，发表于三大顶级经济学期刊：《美国经济评论》（American Economic Review）、《经济学季刊》（Quarterly Journal of Economics）和《政治经济学杂志》（Journal of Political Economy）上的共计5万多个统计检验及其作者的个人资料作为数据来源，并绘制出了这些检验的z统计量（即标准化之后的p值）分布图（见封面图）。

从图中我们可以看到，这些统计检验的z值呈现出了令人吃惊的双峰分布：有大量检验得出了不显著的结果；随着其z值增加，得到发表的检验数量逐渐减少；但在z值达到1.96，即5%显著性水平这一门槛附近时，又开始有越来越多的检验被发表在三大经济学顶级刊物上。在考虑了四舍五入导致的误差，并对包含检验数目不同的论文进行了权重调整之后，我们仍然可以观察到上述的现象。

按照学界的惯例，5%及更高的显著性水平被视为较为合意的结果。因此，z值分布在1.96附近的反常，不由让人怀疑这些检验结果遭遇了人为的操纵。不过，会不会是经济学期刊对于显著结果的偏爱才导致了双峰分布的出现呢？四位研究者指出，如果真是这样，那我们只会发现得到发表的检验数量随着z值增加而不断上升，并不会观察到图1中先下降再上升的趋势。

为了精确估计期刊偏好所产生的影响，研究者们假定，一篇实证文章能否被期刊接受，取决于两方面因素：结果的统计显著性，以及其他方面（题材的重要程度，方法的原创性等等）的质量；在其他方面的质量相同时，期刊会愿意发表结果更显著的论文；如果不存在人为因素影响，那么论文在其他方面是否优秀的概率，应该与其结果的统计显著性互相独立。在此基础上，给定全部投稿论文的先验质量，我们就能估算出理想状态下，期刊上统计检验的z值应该表现出来的分布。先前我们观察到的实际分布减去这一理想分布之后剩下的残差，就是统计检验的z值变化中，无法由期刊偏好解释的部分。

研究者们发现，上述残差在z值达到1.96之前几乎为零，在那之后则迅速上升。对于不同的投稿论文先验质量分布和不同的研究子领域，这一结果都是稳健的。以不显著结果作为主要贡献的论文由于数量太少，也不可能是造成双峰分布的原因。

更为有趣的是，那些未用星星或粗体字标识显著结果的论文——可能意味着作者不太在乎结果的统计显著性——表现出了更小的残差。如果论文是基于某个理论模型或者随机实验方法，其残差会明显低于平均水平；稳健性检验的结果则往往会有高于主要检验的残差。最后，相比于青年研究人员，较为资深、拥有终身教职的学者发表的论文残差也会更小一些。总的来说，在那些研究者有激励，或者更容易操纵统计结果的情况下，我们会更频繁地观察到z值不能被期刊偏好解释的异常变动。研究者的计算显示，那些有人为操纵迹象的结果，占了全部边际显著（marginally significant，即z值刚刚超过1.96）检验的10%-20%。

当然，统计趋势并不能作为人为操纵确实存在的决定性证据。四位研究者所依赖的研究假设也远非无懈可击。然而，这一研究至少说明，人们对统计显著性通胀的批评，似乎不能仅仅看作空穴来风。去年政治学中出现的Lacour丑闻（见政见之前的文章《图文详解：<科学>杂志论文数据造假现形记》），以及心理学中重复经典实验的一再失败，迫使学者们开始思考这样的问题：我们得到的“结论”到底是对世界的真实认知，还是用统计学小把戏变出的障眼法？在统计显著性的诱惑面前，社会科学家的节操还能坚挺多久？

幸运的是，在各个社科领域之中，总还有一些节操高于平均水平的成员，在不断地默默推动着本学科在严谨性方面的进步。本文反复提及的四位研究者本身都是经济学家，心理学中的重复实验也是由心理学家倡导发起，Lacour 丑闻也是由政治学家David Broockman 最先揭发。为了限制对统计结果的人为操纵，新的方法和标准也在不断被学界提出。目前在经济学和政治学的主流期刊上，公布原始数据已经成了一个基本的发表要求。很多经济学家和政治学家甚至会主动公布实证研究使用的软件代码（比如无辜中枪的阿西莫格鲁老湿）。伯克利的经济学家 Ted Miguel，近年来不断和同事一起，倡导“Pre-Analysis Plans”的广泛应用，即在正式开始研究之前，学者先把既定的研究方案公布在网上，以避免事后修改初始目标或方法之类的事情出现。

由于学术界的竞争日益激烈，显著性发生通货膨胀其实并不算出人意料。但是，我们不应因此而动摇对经济学，甚至全部社会科学研究的信心。毕竟，有人为操纵嫌疑的结果总是少数。随着定量技术的进一步完善，大数据分析的逐渐普及，和学术研究透明化程度的不断提高，我们有理由期待，未来的实证研究可以带给人们更加客观，更加确定，更加严格的结果。

参考文献：

Brodeur, Abel, Mathias L., Marc Sangnier, and Yanos Zylberberg. 2016. "Star Wars: The Empirics Strike

财经新闻 项目交易 求职招聘 在线教育 投稿 服务一览 快速导航

1 提问 悬赏 求职 新闻 读书	Casey, Katherine, Rachel Glennerster, and Edward Miguel. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan*." The Quarterly Journal of Economics 127.4 (2012): 1755-1812.
2 经济学人	
3 经济学论坛	
4 工商管理论坛	Masicampo, E.J., and Daniel R. Lalande. "A peculiar prevalence of p values just below .05." The Quarterly Journal of Experimental Psychology 65.11 (2012): 2271-2279.
5 计量经济学与统计论坛	
6 金融投资论坛	Miguel, E., et al. "Promoting Transparency in Social Science Research." Science 343.6166 (2014): 30-31.
7 会计与财务管理论坛	来源：知乎 www.zhihu.com
8 世界经济与国际贸易	作者：王也 http://www.zhihu.com/people/wang-ye-80
9 考研&考博&留学	微信原文：http://mp.weixin.qq.com/s?__biz=MzA3NDkyNTc4Ng==&mid=402255130&idx=2&sn=f81a75417e60d9c74a96f82690143597&scene=4
10 网络课堂	
11 站务区	
12 休闲区	

学习经管知识 掌握财经技能

释疑解惑 知识成就

论坛好贴推荐



CDA数据分析师系列丛书10月又出三本新作！

- 2015年度论坛精彩大事盘点巡礼
- 2015年论坛自制图书！见者有份
- 2015年度论坛30大原创好贴集锦
- 2015年度论坛30大版主精彩发贴
- 2015年度论坛30大经管资源珍藏
- 2015年度论坛30大精彩专题汇总
- 2015年度论坛30大热点人气话题
- 2015年度计量统计40问经典答疑

查看 >

关闭



长按指纹 → 识别二维码 → 一键关注

有经管问题 上人大经济论坛

分享

收藏 6

支持 0

反对 0

相关帖子

- 求 北京市经济社会统计报告2012
- 统计专业的本科生都学哪些课程
- 统计
- 十一五期间江西省统计年鉴
- 1998-2011年中国分地区造林面积统计
- 求 2013年1-3月重庆市保险行业情况统计
- 我国科技统计数据2012
- 中国区域经济年鉴2005 可复制版本

支持楼主：购买VIP、购买贵宾

购买后，论坛将把您花费的资金全部奖励给楼主，以表示您对TA发好贴的支持

相关链接：

经管考试频道----国际内审师备考辅导



论坛客户端，下载免流量

关注论坛官方微信



已有 **1** 人评价 经验 理由 收起

 remlus + 100 精彩帖子

总评分: 经验 + 100 [查看全部评分](#)

[经济学家](#), [统计](#)

[使用道具](#) [举报](#)



[谁边](#) 发表于 2016-3-2 22:05:47 | [只看作者](#) 沙发

Acemoglu 不是主要做理论模型的吗？很少见他的应用文章。

[使用道具](#) [举报](#)



[hdfhj](#) 发表于 2016-3-2 22:16:15 | [只看作者](#) 藤椅

我收集了1949-2015年的数据，为了得出我想要的结论，我不断地play with the data, 最后发现1970-1998年这一时间段能得到“显著”的结果，于是就用这段数据写成论文，全然不提其它时间段的failure.

我这种做法属于“统计学小把戏”，学术不端, 或是什么其它罪名？

[使用道具](#) [举报](#)



[hdfhj](#) 发表于 2016-3-2 22:26:32 | [只看作者](#) 板凳

我一直有个疑问：怎么判断经济学实验的数据是不是伪造的？伪造自然科学实验的数据有风险，因为别人如果重复你设计的实验而得不出你所说的结论，那你就被逮住了。事实上，自然科学实验数据造假被戳穿的事情，国内外经常有发生。

伪造经济学实验数据，显然没有风险。从实验经济学诞生之日起，也没有听说谁因为伪造经济学实验的数据而被抓的。

那么，我们凭什么相信经济学实验数据的真实性呢？

[使用道具](#) [举报](#)



[终结天狼](#) ▼ 发表于 2016-3-2 22:39:20 来自手机 | [只看作者](#) 报纸

[hdfhj](#) 发表于 2016-3-2 22:26

我一直有个疑问：怎么判断经济学实验的数据是不是伪造的？伪造自然科学实验的数据有风险，因为别人如果重复 ...

经济学的数据大多不是实验出来的，即便是实验出来的，只要论据和论证过程没有问题，那么结论应该也不会有问题

[使用道具](#) [举报](#)

[谁边](#) 发表于 2016-3-2 23:16:32 | [只看作者](#) 地板



hdfhlj 发表于 2016-3-2 22:26

我一直有个疑问：怎么判断经济学实验的数据是不是伪造的？伪造自然科学实验的数据有风险，因为别人如果重复 ...

最近参加了两个实验的pilot。说句老实话，经济学实验根本不需要伪造数据的。

对于那些有方向性的实验，例如关于信息不对称下的学习过程，或者配对实验，只要你实验设计得好，你就一定能得到你想要的结果。

对于那些没有方向性的，比如测算warm glow，测算时间折扣，不论你得什么结果，都可以以此写文章，也没有伪造的必要。

但是我读到过一些社会心理学实验造假的新闻。

使用道具 举报



hdfhlj 发表于 昨天 00:05 | 只看作者

7楼

谁边 发表于 2016-3-2 23:16

最近参加了两个实验的pilot。说句老实话，经济学实验根本不需要伪造数据的。

对于那些有方向性的实验，例 ...

按你的逻辑，自然科学家也没必要实验造假了，因为他们面临的情况也差不多（有方向性/没有方向性）。那么为什么还有这么多自然科学家的实验造假呢？

使用道具 举报



谁边 发表于 昨天 01:47 | 只看作者

8楼

二者面临的情况当然不同。评价标准宽严就不同。这也是为什么实验在经济学中的地位远不如它们在自然科学中的地位高。

我不想诋毁从事实验经济学的经济学家。他们的工作可以提供很多有趣的研究方向，有时候也会颠覆一些传统观念，给人耳目一新的感觉。但实验结果对于理论的验证作用不是那么明显。对参数的取值，也很少有经济学家使用实验结果，而大多使用对现实数据的回归结果或者通过校准模型来得到。

使用道具 举报



hdfhlj 发表于 昨天 10:04 | 只看作者

9楼

我收集了1949-2015年的数据，为了得出我想要的结论，我不断地play with the data，最后发现1970-1998年这一时间段能得到“显著”的结果，于是就用这段数据写成论文，全然不提其它时间段的failure.

我这种做法属于“统计学小把戏”，学术不端，或是什么其它罪名？

使用道具 举报



很大的小 发表于 昨天 19:39 | 只看作者

10楼

计量中的确有太多张五常说的“花招”！但是，没办法，因为计量是整个社会科学用来验证因果关系（不是因果逻辑）的唯一方法（不看好经济学中的实验）。

使用道具 举报

 [发帖](#)


1

2

[下一页](#)

[返回列表](#)

[高级模式](#)

您需要登录后才可以回帖 [登录](#) | [我要注册](#)  [用QQ帐号登录](#)

[发表回复](#)

[回帖后跳转到最后一页](#)



[手机版](#) | [意见反馈](#) | [帮助](#) | [新手入门](#) | [用户手册](#) | [友情链接](#) |

4000-51-9191



[联系我们](#)

[招聘英才](#)

[合作加盟](#)

[广告服务](#)

为做大做强论坛，本站接受风险投资商咨询，请联系 (010-62719935)

联系QQ : 75102711 邮箱 : service@pinggu.org

合作咨询电话 : (010)62719935 广告合作电话 : 13661292478 (刘老师)

投诉电话 : (010)68466864 不良信息处理电话 : (010)68466864

京ICP备11001960号 京ICP证090565号 京公网安备1101084107号 论坛法律顾问：王进律师 知识产权保护声明 免责及隐私声明