

社会学教材教参方法系列

SM

年龄-时期- 队列模型 ——聚合数据分析方法

Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data

〔美〕罗伯特·M. 奥布莱恩 (Robert M. O'Brien) / 著

王培刚 姜俊丰 等/译

海外借



社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

SM

年龄-时期- 队列模型

——聚合数据分析方法

**Age-Period-Cohort Models: Approaches and
Analyses with Aggregate Data**

[美] 罗伯特·M. 奥布莱恩 (Robert M. O'Brien) / 著

姜俊丰 等/译



图书在版编目(CIP)数据

年龄-时期-队列模型: 聚合数据分析方法 / (美)

罗伯特·M. 奥布莱恩 (Robert M. O'Brien) 著; 王培刚等译. -- 北京: 社会科学文献出版社, 2018. 10

(社会学教材教参方法系列)

书名原文: Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data

ISBN 978-7-5201-2957-2

I. ①年… II. ①罗… ②王… III. ①统计数据-统计分析 IV. ①O212.1

中国版本图书馆 CIP 数据核字 (2018) 第 134083 号

社会学教材教参方法系列

年龄-时期-队列模型

——聚合数据分析方法

著 者 / [美] 罗伯特·M. 奥布莱恩 (Robert M. O'Brien)

译 者 / 王培刚 姜俊丰 等

出 版 人 / 谢寿光

项目统筹 / 杨桂凤

责任编辑 / 杨桂凤 杨鑫磊

出 版 / 社会科学文献出版社·社会学出版中心 (010) 59367159

地址: 北京市北三环中路甲 29 号院华龙大厦 邮编: 100029

网址: www.ssap.com.cn

发 行 / 市场营销中心 (010) 59367081 59367018

印 装 / 三河市龙林印务有限公司

规 格 / 开 本: 787mm × 1092mm 1/16

印 张: 13 字 数: 229 千字

版 次 / 2018 年 10 月第 1 版 2018 年 10 月第 1 次印刷

书 号 / ISBN 978-7-5201-2957-2

著作权合同 / 图字 01-2017-4966 号

登 记 号

定 价 / 69.00 元

本书如有印装质量问题, 请与读者服务中心 (010-59367028) 联系

 版权所有 翻印必究



主编◎谢宇 郭志刚

Age-Period-Cohort Models:

Approaches and Analyses with Aggregate Data

By Robert M. O'Brien / ISBN 978 - 1 - 4665 - 5153 - 4

© 2015 by Taylor & Francis Group, LLC

All Rights Reserved

Authorized translation from English language edition published by Chapman and Hall/CRC Press, an imprint of the Taylor & Francis Group LLC. All Rights Reserved. 本书原版由 Taylor & Francis 出版集团旗下 Chapman and Hall/CRC 出版公司出版，并经其授权翻译出版。版权所有，侵权必究。

Social Sciences Academic Press (China) is authorized to publish and distribute exclusively the Chinese (Simplified Characters) language edition. This edition is authorized for sale throughout Mainland of China. No part of the publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher. 本书中文简体翻译版授权由社会科学文献出版社独家出版并限在中国大陆地区销售，未经出版者书面许可，不得以任何方式复制或发行本书的任何部分。

Copies of this book sold without a Taylor & Francis sticker on the cover are unauthorized and illegal. 本书封面贴有 Taylor & Francis 公司防伪标签，无标签者不得销售。

前 言

年龄效应、时期效应以及队列效应的分离是所有涉及一个或多个这些效应的分析的关键问题，但这个问题仍然常常被忽视。在横截面调查中，研究者可能会发现吸烟行为与年龄有关。这一研究通过检验单一时期的数据从而控制了时期效应，使时期的方差为零，因此其“效应”为零。然而，这个横截面研究中“显而易见的”“年龄效应”也有可能是队列效应，或者是年龄效应与队列效应的混合，因为在这个横截面研究设计中，每个年龄组都代表一个不同的出生队列。在检验不同时期中的同一年龄组，或者不同年龄组中的同一队列时，类似的问题也会存在。这些情况均涉及被遗漏的变量。

研究人员经常会在其研究设计中考虑这三个变量中的两个。研究人员一开始可能设计了年龄-时期研究，关注在不同时期中年龄分布如何波动，继而解释为什么时期影响了结核病或同性婚姻支持度或自杀的年龄分布。使用这种研究设计的研究者很可能忽略了可能存在的队列效应。但队列效应不应被忽视，因为年龄、时期和队列效应是相混杂的，已知年龄和时期，我们就可得知出生队列。这些自变量之间存在线性依赖关系（时期-年龄=队列）。

在上述情况下，研究者可能没有意识到他们面临着年龄-时期-队列（APC）识别问题，然而多数情况下事实确实如此。任何在分析数据时明确考虑到年龄、时期和队列效应的研究者几乎都认识到在此模型中的识别问题，他们认识到了包含年龄、时期和队列效应的模型会存在线性依赖关系。

本书是对利用聚合层次数据建模分析年龄、时期和队列效应的问题及策略的

介绍。这些策略涉及约束估计、因素 - 特征法、可估函数以及方差分解。第 1 章为综合且全面的概述性章节，之后的第 2 ~ 3 章分别从代数和几何角度说明了识别问题，其中也讨论了约束回归。第 4 ~ 6 章提供了一些不直接依赖第 2 ~ 3 章中的约束条件来识别 APC 模型的方法。第 7 章展示的一个具体实证案例表明，多种方法的结合使用可以为特定的年龄、时期和队列效应提供更有说服力的证据。

关于 APC 模型的文献浩如烟海。在本书中，我通过对使用聚合层次数据的 APC 模型进行检验，从经典视角对其中的部分文献进行介绍。这些数据来自公开的《人口统计》、《统一犯罪报告》和其他报告了不同年份中分年龄组的凶杀案例数、结婚数或结核病死亡数的政府报告。同类的分析经常基于在不同时期内进行的重复横截面调查的数据而展开，数据聚合后可得到年龄分布。虽然有一些 APC 模型的方法同时使用个体层次和聚合层次数据来估计年龄、时期和队列效应，但我将重点放在了应用聚合层次数据的问题和策略上。在第 5 章的方差分解法中，我对此类混合层次模型做了评论。第 1 章结尾对全书的设计及结构做了介绍。

我需要感谢一些重要人物。威廉·比尔·马松 [Willian (Bill) Mazon] 于 20 世纪 80 年代末在俄勒冈大学发表了关于政治疏离和相对队列规模的演讲，这使我对 APC 视角产生了兴趣。他的演讲十分清晰，以至于我都可以重现其中谈到的分析。珍·斯托卡德 (Jean Stockard) 对我给她展示的青少年凶杀犯罪流行特征图表给出了一个可能的解释，这让我重新燃起了对 APC 分析的兴趣。我们二人有着最丰富的合作发表文章的经历。克里斯·温希普 (Chris Winship) 作为《社会学方法与研究》(*Sociological Methods and Research*) 杂志的主编，一直与我探讨我在在审文章中所主张的内容，与我探讨如何更好地解释这些内容。他也督促我和对我的文章做出评价的同行们以礼相待，讨论问题时对事不对人。我自己作为一本社会学杂志《社会学视角》(*Sociological Perspectives*) 的合作主编，唯有向着达到上述品质的目标而努力。除此之外，许多有天赋的老师在我于波莫纳学院读本科时将我“引向正路”。当我刚进入大学时，我更关注篮球而不是学术。威斯康星大学的教授们也花时间将我培养成了专业的社会学者，当然，也花时间灌输了传统优良的中西部经验主义价值。我的大儿子乔西·奥布莱恩 (Josh O'Brien) 和克里斯·温希普 (Chris Winship) 在我还没想到时就建议我写一本书，于是我联系了查普曼与霍尔出版公司 (Chapman & Hall)，余下的事情都如你们所知了。

目录

CONTENTS

前 言	1
1 年龄 - 时期 - 队列模型概述	1
1.1 引言	1
1.2 对年龄、时期和队列的关注	2
1.3 队列的重要性	5
1.4 本书的计划	16
参考文献	18
2 多分类模型和约束回归	21
2.1 引言	21
2.2 线性编码的年龄 - 时期 - 队列 (APC) 模型	22
2.3 分类编码的 APC 模型	24
2.4 广义线性模型	29
2.5 零向量	29
2.6 模型拟合	31
2.7 解与约束条件正交	32
2.8 检验解之间的关系	32
2.9 约束解在旋转前后的差异	40
2.10 忽略一个或多个年龄、时期或队列因素的解	42
2.11 偏差: 约束估计和数据生成参数	46
2.12 单一约束条件下的无偏估计	48

2.13	有额外实证支持的合理约束条件	48
2.14	结论	51
附录 2.1	虚拟变量和效应编码	52
附录 2.2	确定效应编码、虚拟变量编码变量的零向量	53
附录 2.3	作为无偏估计的约束估计	55
	参考文献	55
3	APC 模型和约束估计的几何原理	57
3.1	引言	57
3.2	一般几何视角下的单一秩亏模型	58
3.3	多维模型的泛化	65
3.4	含线性编码变量的 APC 模型	66
3.5	几何解和代数解的等价性	73
3.6	多分类模型的几何原理	75
3.7	离原点的距离和沿解集线的距离	76
3.8	实证案例：Frost 的结核病数据	77
3.9	APC 模型几何原理重要特征总结	80
3.10	机械约束的问题	82
3.11	讨论	85
	附录 3.1	85
	参考文献	86
4	可估函数法	88
4.1	引言	88
4.2	可估函数	89
4.3	在年龄 - 时期 - 队列 (APC) 模型中用 $l'sv$ 法建立可估函数	91
4.4	利用 $l'sv$ 法推导可估函数的一些示例	93
4.5	对 $l'sv$ 法的评论	99
4.6	有经验数据的可估函数	100

4.7 对男女肺癌死亡率差异的更多实质性检验	104
4.8 结论	106
附录 4.1	107
参考文献	109
5 在年龄-时期-队列 (APC) 模型中分解方差	111
5.1 引言	111
5.2 归因方差: 年龄-时期-队列方差分析法 (APC ANOVA)	112
5.3 APC 混合模型	115
5.4 分层 APC 模型	124
5.5 使用凶杀犯罪数据的实证案例	127
5.6 结论	136
参考文献	138
6 因素-特征法	140
6.1 引言	140
6.2 单因素特征	141
6.3 双因素或多因素的特征	145
6.4 因素和因素特征的方差分解	146
6.5 实证案例: 年龄-时期别自杀率和频数	147
6.6 对具有两个队列特征的自杀数据进行年龄-时期-队列特征 (APCC) 分析	152
6.7 对具有两个时期特征的自杀数据进行年龄-队列-时期特征 (ACPC) 分析	155
6.8 年龄-时期-特征-队列特征模型	158
6.9 基于因素特征和机制的方法	160
6.10 因素-特征模型的其他特征和分析	161
6.11 结论	162
参考文献	163

7 总结：一个实证案例	165
7.1 引言	165
7.2 实证案例：凶杀犯罪	166
7.3 结论	183
参考文献	185
索引	187
译后记	193



年龄 - 时期 - 队列模型概述

正如掌握细菌学的基础知识一样，了解流行病学史并牢固掌握相关统计方法是流行病学研究人员的基本素养。

Major Greenwood (Hardy and Magnello, 2004: 214)

1.1 引言

本书开篇引文表明，流行病学研究人员需要对诸如流行病学、人口学或某一社会科学以及对应的统计方法等多个领域有实质性认识。更明确地说，对于本书而言，研究人员需要具备大量知识，包括年龄、时期和队列对某个领域内的某个真实变量的潜在影响，以及用于研究年龄、时期和队列之间关系的统计学方法。笔者不会介绍任何一个领域的实际知识或是综述某一领域中年龄 - 时期 - 队列模型的文献，笔者另有目的。

本书的重点在于深入了解与年龄、时期和队列分析相关的统计条件和统计难题，并在此基础上评估在文献中应用最广泛的几种方法之间的相互联系，以及如何应用这些方法找出年龄、时期和队列因素与真实变量之间的关系。深入理解这些后，读者便可以评估文献中出现的新方法并充分利用现有的方法。读者也可以用严谨、稳健的方法来审慎地解释这些分析的结果。

本章介绍了流行病学家、人口学家和社会科学家在研究年龄、时期和队列与关键因变量之间的联系时所关注的领域，并援引了实际领域中的一些案例予以说

明。由于这些研究领域均涉及 3 个相同的自变量（年龄、时期和队列），这些实际领域的情况也包括诸如生育率、结核病的死亡率、自杀率、同性婚姻支持度或政府信任度之类的结果变量。^①

1.2 对年龄、时期和队列的关注

从某种角度来说，年龄 - 时期 - 队列（age-period-cohort, APC）难题似乎只是计量方法专家应该关注的一个深奥问题。从另一种角度来说，这个难题是人口学、经济学、流行病学、政治学、社会学和相关领域内大量研究的核心问题。该观点似乎与实证文献相悖。有多少研究者在他们的研究中试图区分年龄、时期和队列？遗憾的是，这种情况并不多见。然而我们常常发现，研究者在横截面分析中会讨论年龄因素，或者探讨年轻人或者老年人的行为模式随时期的变化。研究者也会在某个单一模型中采用两个变量，例如研究疾病或自杀的年龄分布，或人们的态度在不同时期中的变化。在这些案例中，APC 识别问题经常潜藏在研究者没有考虑到的地方。

1.2.1 年龄模型

如果研究者在开展于 1970 年的横截面研究中发现肺癌死亡率的峰值分布在 45 ~ 54 岁这一年龄段会怎样？这个有趣的发现或许表明处于某一年龄组的群体更可能死于肺癌，并为肺癌的病因学研究提供线索。从表面上看，这个发现似乎与 APC 难题没有什么关联，但仔细观察会发现它与 APC 难题密切相关。这种年龄分布很可能是吸烟量的队列差异所导致的。在 1970 年处于 45 ~ 54 岁的队列成员在 1940 年则处于 15 ~ 24 岁（吸烟习惯大致养成的年龄段）。如果在 1940 年年龄在 15 ~ 24 岁的吸烟者数量非常多（比前后队列中的吸烟者数量都多），那么（在其他条件均一致的情况下）当该队列成员的年龄在 45 ~ 54 岁时，其肺癌死亡率将高于其他年龄组。仅单独研究年龄效应并不能解决 APC 难题，因为这一做法忽略了潜在的队列效应。这种横截面研究中的年龄分布可

① 本书通篇集中关注重复横截面数据，而非在一定时间内对固定的调查对象进行调查而获得的追踪数据。本书假设研究对象不会由于移民或其他因素而出现重大变化，以至于影响因变量和年龄、时期、队列因素之间的关系。

能仅仅是年龄效应导致的,也可能仅仅是队列效应导致的,或者是年龄和队列效应共同导致的。

1.2.2 时期模型

在研究时间序列时,研究者可能会发现,在全国民意调查中,30~39岁人群对同性婚姻的支持度有上升的趋势,并据此得出结论:随着时间的推移,人们越来越支持同性婚姻。该时期效应可以解释为:人们的观点往往随着时间的推移而逐渐发生变化。但这同样可以解释为队列效应,即相比较早期的队列,晚近队列的成员更加支持同性婚姻。也就是说,促使人们对同性婚姻支持度发生变化的不是人们态度的转变,而是因为同一个年龄组成员的态度代表着晚近时期中晚近队列成员的态度,即队列效应。同样,仅单独研究时期效应并不能解决APC难题,因为这一做法忽略了一个解释的另一方面。在该时间序列研究中,人们对同性婚姻支持度的变化可能仅仅是时期效应导致的,或仅仅是队列效应导致,也可能是时期和队列效应共同导致的。

1.2.3 队列模型

单独分析队列效应的研究更为罕见。研究者可以通过比较处于同一个年龄组(40~49岁)的不同队列来确定来自其中某些队列的这一中年选民群体对社会福利项目的支持度是否更高。如果队列之间存在系统的差异,这是队列效应还是时期效应造成的?每个队列中40~49岁年龄组的研究个体均处于不同的时期。这些明显的队列差异可能仅仅是队列效应或时期效应造成的,也可能是队列和时期效应共同造成的。

如前所述,仅考虑年龄因素的设计通过考察某个特定时期的年龄因素来控制时期效应,仅考虑时期因素的设计通过考察某个特定的年龄组来控制年龄效应,而仅考虑队列因素的设计也通过考察某个特定的年龄组来控制年龄效应。这些模型设计出现误设的可能性都非常高(两个变量之间的关系可能完全或部分依赖于已经被排除在模型之外的另一个变量)。即使我们将其中的两个变量(如年龄和时期、年龄和队列或时期和队列)纳入研究中,这种误设的风险仍然存在。例如,如果笔者在一般性的研究设计中研究一些重要因变量(如随时间变化的凶杀率、结核病死亡率或同性婚姻支持度)的年龄分布,那么问题在于:随着时间的推移,年龄效应是否能够保持稳定?

1.2.4 年龄 - 时期模型

在犯罪学中,一些研究者认为凶杀犯罪和其他一些犯罪行为的年龄分布是不变的(Hirschi and Gottfredson, 1983),但其他一些研究者却对这个观点提出了质疑(Greenberg, 1985; Steffensmeier, Allan, Harer, and Streifel, 1989)。在20世纪80年代中期,凶杀犯罪的年龄分布发生了巨大的变化,15~19岁年龄组的凶杀率上升了1倍多,20~24岁和25~29岁年龄组的凶杀率也有所上升,但其他更老年龄组的凶杀率则有所下降。不同年龄组在不同时期内凶杀犯罪倾向性的变化是否可以解释为一种具有稳定年龄效应的新型凶杀犯罪模式?这种变化是不是一种与时期变化密切相关的持续性变化,如年轻人的角色变化,针对年轻人犯罪的刑事司法政策的变化,或其他与凶杀犯罪年龄分布有关的时期变化?答案或许是肯定的。然而,由于每个时期中的不同队列均代表了某个年龄组,年龄分布在不同时期的变化也可能归因于队列效应,而在这一案例中,这种队列效应与凶杀犯罪的年龄分布是相互混淆的。每个时期的年龄分布都不仅仅是年龄和时期的综合作用,在APC背景下,我们必须考虑到队列的潜在作用(O'Brien, Stockard, and Isaacson, 1999)。

1.2.5 年龄 - 队列模型

研究者可能会采用年龄 - 队列模型来研究年龄效应,探究不同队列中的因变量与年龄之间的关系是否存在一些相似性或差异性。该因变量可能是年龄 - 队列别政治保守化程度得分、同性婚姻支持度得分或癌症死亡率。研究者可能会发现,因变量随年龄增长的模式在几个队列中是相同的。如果不同队列的年龄模式是相似的,那么似乎可以认定这种模式是由特定的年龄效应造成的。当然,这只是一种可能性,但这种变化模式也可能归因于时期效应。如果线性时期效应随着因变量的年龄变化而出现增减,那么便很可能出现这种混淆。其他的变化模式也十分令人费解。更具体地说,如果同性婚姻支持度是一种时期效应,即随着时间的推移,人们普遍倾向于支持同性婚姻,那么,同性婚姻支持度也可能被视为队列中的一种年龄效应,即随着年龄的增长,同性婚姻支持度提高了。同性婚姻支持度更有可能同时受到年龄效应和队列效应的影响,区分这些效应是APC难题中的一部分。

1.2.6 年龄 - 时期 - 队列模型

鉴于仅仅分析年龄、时期、队列因素中的某一个或任意两个因素时可能会出现混淆,研究者可能希望在标准模型中合理地识别这三个变量的作用。然而,这种标准模型是用于分析某一个或任意两个因素的。由于这三个变量间存在线性依赖关系,同时分析这三个因素会导致无法区分年龄、时期和队列效应。这三个因素中的任意一个均不能完全独立于另外两个因素。在最简单的情况下,即按年份对年龄、时期和队列进行线性编码,如果我们知道某一类人所属的时期和年龄,那就可以确定他们的出生队列:时期 - 年龄 = 队列。那么问题就在于:因变量与队列的关系是与队列效应有关还是与时期效应减去年龄效应有关?除非对这三者之间的关系做出一个假设,否则我们无法得到答案。这正是本书所关注的重点。笔者致力于研究这个难题,并探讨如何在只能获得聚合层次数据的情况下从这些模型中获得信息。政府机构提供的大量数据属于聚合层次数据,这种数据也是APC模型的基本关注点。

1.3 队列的重要性

与在大量社会和行为科学研究中发挥重要作用的年龄组和时期相比,队列在现有文献中很少受到关注。^① 公众认知也同样存在此种倾向,人们更容易理解年龄在凶杀率、出生率、政治保守度或结核病死亡率和自杀率随时间的推移发生改变中所起的作用。随之而来的问题通常是年龄分布的模式是什么及其如何随时间而变化。当提及队列或出生队列在结核病死亡率或凶杀率中发挥的作用时,人们可能会问:什么是队列?即使解释了什么是队列,人们也很难理解队列和凶杀率或结核病死亡率之间的关系。人们似乎不太适应或至少不习惯从队列的角度考虑问题。然而,当队列差异(cohort differences)表现为世代差异(generational differences)(例如生育高峰和生育低谷世代)时,人们更容易理解这一观点。对于非专业人士来说,成年于大萧条时代的那代人在经济上更节俭或政治上更倾向于自由主义这种说法似乎更加容易理解。

① 这并不是说人口学家、流行病学家及社会科学家忽略了队列因素(远非如此),但是在很多横截面研究中,只有年龄才是一个标准变量,在时间序列分析中则是时期变量。队列在研究中较少受到关注。

在本章中，笔者回顾了许多广受认可的高质量论著，这些论著采用不同的方式研究了年龄、时期和队列的一般问题。其中，有几篇来自不同学科的里程碑式论文成为鼓励各学科加大对队列因素的研究，以及对由年龄组、时期和队列因素的混杂效应所产生的难题进行研究的试金石。Norman Ryder (1965) 曾在人口学和社会学领域发表了一篇名为《队列是社会变革研究中的重要概念》的里程碑式论著，这篇论著及这两个学科都着重提到了 Karl Mannheim (1928/1952) 的论著《世代的问题》。Mannheim 则在其研究中引用了 Hume 和 Comte 的论点。尽管这些文章都没有重点分析实证数据，但它们都呼吁研究者关注队列及其在社会变迁和稳定中的作用。列克西斯 (Lexis) 图可追溯到 19 世纪 60 年代或 70 年代^① (将在本章后面进行讨论)，尽管列克西斯图的核心要义是将出生、年龄和时期这三个时间维度联系在一起，但它并没有揭示队列的重要性。在流行病学领域内，Wade Hampton Frost (1939) 撰写了一篇涉及实证数据检验的奠基之作，这篇论著在他于 1938 年去世不久后出版。寿命表是人口学、流行病学和精算科学的核心。在最初的发展中，寿命表的计算并未考虑队列效应，但现在队列通常在寿命表的计算中发挥着重要作用。尽管本书可以列举数十种有关 APC 模型类型的先驱性研究和分析成果，但正如本章的第一段所述，本书的目标并不在于探究年龄、时期和队列分析的发展历史。尽管介绍年龄和时期效应的相关研究看起来似乎更加“合乎常理”，笔者还是仅援引上述论著对队列效应的发展情况进行一些介绍。

1.3.1 寿命表

寿命表在人口统计学和流行病学中发挥着重要作用（在犯罪学、经济学、社会学等学科中也发挥着作用）。John Graunt (1662) 被公认为是寿命表制作的鼻祖。他指出：“然而，我们发现 100 例活产儿中约有 36 例死于 6 岁之前，可能有 1 例存活到 76 岁。6 ~ 76 岁有 70 年，我们从存活的 64 个人（6 岁时存活，最后一位存活到 76 岁）中找出 6 个比例中项 (mean proportional numbers) ……”需要注意的是，Graunt 仅纳入了活产儿，所以出生时的死亡人数为零。同时他也需要“估计”6 ~ 76 岁每 10 年间的死亡人数，数据结果详见表 1 - 1。由于

^① 尽管人们质疑该图是由经济学家 Lexis 于 1874 年编制的还是由 Zeuner 在 1869 年或 Brasche 在 1870 年编制 (Vandeschrick, 2001) 的，但该图表依然用 Lexis 来命名。