

老乡见老乡 深圳市不同方言群体的空间聚集、行业聚集与收入回报

普娟菠 曾东林 张卓妮*

摘要:移民群体的空间聚集和行业聚集是城市研究的重要议题。与以往研究不同,本文以“方言文化差异”的视角研究我国流动人口的内部差异。我们利用深圳市 2005 年全国 1%人口抽样调查数据,比较客家、粤、赣、潮州、湘和闽等六大方言群体的居住聚集、行业聚集与收入之间的关系,以检验空间同化理论和分层同化理论对中国城市不同方言群体的适用性。我们发现,对客家、粤和闽群体而言,居住聚集和收入呈负相关,那些在所属群体聚集的社区居住的人,相比在非聚集区居住者,平均月收入更低;但对赣和潮州群体而言,居住聚集和收入呈正相关:说明空间同化理论的适用性因方言文化群体而异。在行业聚集与收入的关系方面,我们的结果跟以往分层同化理论的文献一样,发现同时存在正相关和负相关的情况。对粤群体而言,在所属方言群体聚集行业(相比在非聚集行业)工作的人,平均收入更高;而客家和潮州群体则相反。这些研究结果说明,中国城市移民研究需重视方言文化等因素带来的流动人口内部差异。我们对本文的发现和未来相关研究的方向进行了讨论。

关键词:居住聚集 行业聚集 收入 方言文化差异 空间同化理论 分层同化理论

* 普娟菠(juanbopu2-c@my.cityu.edu.hk),香港科技大学广州校区社会枢纽城市治理与设计学域博士研究生;曾东林(donglin@m.scnu.edu.cn),华南师范大学哲学与社会发展学院特聘研究员;张卓妮(通讯作者,znzhang@ust.hk),香港科技大学广州校区社会枢纽城市治理与设计学域副教授、香港科技大学社会科学部与公共政策学部联属副教授。

一、引言

族群聚集是都市中的普遍现象,根据同质性理论(homophily theory),人们更愿意与同类型的群体接触,并在交往中自我归组。^①这一过程主要包括两种形式:居住聚集与族群经济。居住聚集反映社会群体在空间分布中的隔离状态,而族群经济则强调族群网络在经济关系中的重要作用。西方对族群隔离进行了长期的理论与实证探究:在理论层面,已形成人类生态学派、行为学派、制度学派、新古典经济学派和马克思主义学派等多种学派^②,以及空间同化理论(spatial assimilation theory)和分层同化理论(segmented assimilation theory)等经典理论^③;在实证研究层面,西方已有大量文献对族群隔离的原因及影响展开了探讨,进而对经典理论进行验证。

与西方跨国移民不同,中国国内的迁移人口主要是汉族,因而种族、宗教、国籍等因素并非首要研究议题;但是,流动人口的内部差异尤其是“方言文化差异”突出。有研究表明,中国东部和南部大部分地区都是方言盛行的地区。^④在国内,人们几乎离家 100 公里就能碰到一种不同

① Miller McPherson, Lynn Smith-Lovin, James Cook, “Birds of a Feather: Homophily in Social Networks”, *Annual Review of Sociology*, Vol. 27, No. 1, 2001, pp. 415-444.

② 孙秀林、顾艳霞:《中国大都市外来人口的居住隔离分析——以上海为例》,《东南大学学报》(哲学社会科学版)2017年第4期;钟奕纯、冯健:《城市迁移人口居住空间分异——对深圳市的实证研究》,《地理科学进展》2017年第1期。

③ 彭庆军:《西方城市族群居住隔离的空间整合——理论、政策与反思》,《民族研究》2018年第5期;Zeng Donglin, Wu Xiaogang, Zhang Zhuoni, “Residential and Industrial Enclaves and Labor Market Outcomes among Migrant Workers in Shenzhen, China”, *Journal of Ethnic and Migration Studies*, 2021, pp. 1-23.

④ Gong Yuanyuan, Irene Hau-siu Chow, David Ahlstrom, “Cultural Diversity in China: Dialect, Job Embeddedness, and Turnover”, *Asia Pacific Journal of Management*, Vol. 28, No. 2, 2011, pp. 221-238.

的、难以理解的方言。^① 区域性差异塑造着不同的态度和行为^②，由方言文化而各成群体的“移民”之间是否也存在差异？

以往关于迁移人口居住聚集和族群经济的研究，大多将移民划分为本地人口和流动人口，而忽略了移民内部不同方言群体之间的比较。本研究旨在填补这一空白。深圳是典型的移民城市，大量且多元的流动人口为研究“方言群体”差异创造了条件。我们利用深圳市 2005 年全国 1% 人口抽样调查数据，比较分析客家、粤、赣、潮州、湘和闽等六大方言群体的空间聚集、族群经济与劳动收入的关系。本文的主要贡献在于提出“方言文化差异”的分析视角，进而检验经典的空间同化理论和分层同化理论，其实用性是否在不同的方言群体中存在差别。

二、居住隔离与族裔经济：理论与实证研究回顾

空间同化理论和分层同化理论是族群隔离研究中的两个经典理论。空间同化理论提出，城市中的族群聚集区为“临时庇护所”和“过渡社区”，它们为社会经济地位较低的流动人口提供居住、生活、就业等各方面的庇护。随着该部分群体社会经济地位的提升，他们将搬离聚集区，从而更好地融入主流社会。^③ 对此，很多文献证实了空间聚集与劳动收入之间呈现负相关。^④ 分层同化理论提出一个更广阔的视角，认为不同

① Mary Erbaugh, “Southern Chinese Dialects as a Medium for Reconciliation within Greater China”, *Language in Society*, Vol. 24, No. 1, 1995, pp. 79–94; Lynn Pan, *Sons of the Yellow Emperor: The Story of the Overseas Chinese*, London: Mandarin, 1990.

② Kai-Alexander Schlevogt, “Institutional and Organizational Factors Affecting Effectiveness: Geoeconomic Comparison Between Shanghai and Beijing”, *Asia Pacific Journal of Management*, Vol. 18, No. 4, 2001, pp. 519–551.

③ D. Massey, N. Denton, “Spatial Assimilation as a Socioeconomic Outcome”, *American Sociological Review*, Vol. 50, No. 1, 1985, pp. 94–106.

④ Zeng Donglin, Wu Xiaogang, Zhang Zhuoni, “Residential and Industrial Enclaves and Labor Market Outcomes among Migrant Workers in Shenzhen, China”, *Journal of Ethnic and Migration Studies*, 2021, pp. 1–23.

社会经济背景的移民群体最终会融入不同的社会阶层。^① 双重劳动力市场理论更进一步阐释,认为劳动力可以大致分为高技能劳动力和低技能劳动力,前者受益于人力资本而进入一级劳动力市场,后者则更多流入二级劳动力市场。因而,族裔经济与劳动收入的关系有正相关也有负相关。

对居住聚集与族群经济的理论回应持续受到西方学者的关注。Logan 等学者发现,美国的少数族裔通过建立聚集区为贫困成员提供庇护,以帮助其更好地融入白人社会。^② Massey 和 Denton 用结构方程模型对此进一步证实,发现在美国的西班牙裔及黑人群体中存在着“过渡社区”。^③ Andersen 考察了欧洲的情况,发现在欧洲的少数族裔通过聚集社区来规避住房市场中被歧视的风险。^④ 这些文献均证实了空间同化理论在欧美国家的适用性。对分层同化理论的验证也有类似的结论,Portes 等人的研究发现在迈阿密古巴裔通过族群网络建立起了族裔经济,并实现了超过市场平均水平的劳动回报,同样的现象在纽约市的华人群体中也存在。^⑤

① Carl Bankston, Zhou Min, “The Social Adjustment of Vietnamese American Adolescents: Evidence for a Segmented-Assimilation Approach”, *Social Science Quarterly*, Vol. 78, No. 2, 1997, pp. 508–523; Alejandro Portes, Zhou Min, “The New Second Generation: Segmented Assimilation and Its Variants”, *The Annals of the American Academy of Political and Social Science*, Vol. 530, No. 1, 1993, pp. 74–96.

② John Logan, Richard Alba, Zhang Wenquan, “Immigrant Enclaves and Ethnic Communities in New York and Los Angeles”, *American Sociological Review*, Vol. 67, No. 2, 2002, pp. 299–322.

③ D. Massey, N. Denton, “Spatial Assimilation as a Socioeconomic Outcome”, *American Sociological Review*, Vol. 50, No. 1, 1985, pp. 94–106.

④ Hans Andersen, “Spatial Assimilation in Denmark? Why Do Immigrants Move to and from Multi-ethnic Neighbourhoods”, *Housing Studies*, Vol. 25, No. 3, 2010, pp. 281–300.

⑤ Alejandro Portes, Leif Jensen, “The Enclave and the Entrants: Patterns of Ethnic Enterprise in Miami before and after Mariel”, *American Sociological Review*, Vol. 54, No. 6, 1989, pp. 929–949; Kenneth Wilson, Alejandro Portes, “Immigrant Enclaves”, *The American Journal of Sociology*, Vol. 86, No. 2, 1980, pp. 295–319; Zhou Min, John Logan, “Returns on Human Capital in Ethnic Enclaves: New York City’s Chinatown”, *American Sociological Review*, Vol. 54, No. 5, 1989, pp. 809–820.

国内关于居住隔离与族群经济的研究,以描述性文献为主。随着普查数据和调查数据的开放,近些年逐渐有学者尝试用实证方式对隔离现象进行深入分析。^①张春泥和谢宇利用2010年对珠三角和长三角地区外来务工人员的调查数据,考察了两地同乡聚集经济的情况,发现加入同乡聚集经济的外来务工人员,相比在开放劳动力市场的人,平均收入更高。^②曾东林、吴晓刚和张卓妮分析了深圳市流动人口的居住聚集、行业聚集与就业市场的关系,发现居住于同乡聚集区的群体平均收入更低,但就业于同乡聚集行业的群体平均收入更高。^③这些研究对经典的空间同化理论和分层同化理论进行了回应,推动了国内关于居住隔离和族群经济的实证研究的发展,但是关于移民群体内部差异,尤其是方言群体之间的比较研究方面,存在不足。

语言与族群聚集密切相关。在群体的社交网络方面,学者发现拥有相似语言的群体往往认可相似的文化,也更有可能长期留在同族群网络中。^④对贸易行为的研究也发现了类似的规律,有学者提出较大的语言差异会增加贸易成本,阻碍跨地区的经济来往,但相近的语言能有效促进贸易。^⑤这些研究说明方言文化在群体的居住聚集与族群经济中起着重要作用,不少西方文献也证实了这一点。Gal调查了奥地利的德国

① 孙秀林、顾艳霞:《中国大都市外来人口的居住隔离分析——以上海为例》,《东南大学学报》(哲学社会科学版)2017年第4期。

② Zhang Chunni, Xie Yu, "Ethnic Enclaves Revisited: Effects on Earnings of Migrant Workers in China", *Chinese Journal of Sociology*, Vol. 2, No. 2, 2016, pp. 214 - 234.

③ Zeng Donglin, Wu Xiaogang, Zhang Zhuoni, "Residential and Industrial Enclaves and Labor Market Outcomes among Migrant Workers in Shenzhen, China", *Journal of Ethnic and Migration Studies*, 2021, pp. 1 - 23.

④ Howard Giles, Donald Taylor, Richard Bourhis, "Towards a Theory of Interpersonal Accommodation through Language: Some Canadian Data", *Language in Society*, Vol. 2, No. 2, 1973, pp. 177 - 192; Gong Yuanyuan, Irene Hau-siu Chow, David Ahlstrom, "Cultural Diversity in China: Dialect, Job Embeddedness, and Turnover", *Asia Pacific Journal of Management*, Vol. 28, No. 2, 2011, pp. 221 - 238.

⑤ Wang Ling, Ruan Jianqing, "Dialect, Transaction Cost and Intra-National Trade: Evidence from China", *Applied Economics*, Vol. 51, No. 57, 2019, pp. 6113 - 6126.

和匈牙利社区,发现个人的社会网络与其语言选择密切相关^①, Milroy 对该结论进行了补充,发现英国的语言聚集区在很大程度上也受到社交网络的影响^②。方言文化与社会网络紧密相关,进而影响着族群的居住聚集与族群经济。

在我国,方言文化也是流动人口内部分化的重要因素,所谓“老乡见老乡”,往往包含着明显的群体内部认同以及与此相关的社会网络。因此,我们在这篇文章里提出“方言文化差异”的视角,以深圳为例研究我国“移民”群体的内部差异,进而检验空间同化理论和分层同化理论在不同方言群体中的适用性。

三、基于语言文化的族群内部 差异视角：以深圳为例

深圳市作为中国典型的移民城市,具有多元的人口结构和丰富的方言文化。如图 1 所示,深圳市常住人口的户籍登记地几乎覆盖全国各地区。除户籍地为深圳市的人口之外,来源地为广东省其他地市的人口占比最高,超过 20%;其次为来自湖南、湖北、广西、四川及江西的流动人口,均构成深圳市总人口的 5% 及以上;来源地为河南、贵州、重庆、安徽、陕西及福建的人口同样占据着重要比重,均大于 1%;其余人口的来源地数量多、范围广。据深圳市统计年鉴记载,过去 30 年来流动人口占据着深圳市的绝大部分人口。如此多元的流动人口格局,为“族群”内部异质性的探究提供了重要条件。

① Susan Gal, “Peasant Men Can’t Get Wives: Language Change and Sex Roles in a Bilingual Community”, *Language in Society*, Vol. 7, No. 1, 1978, pp. 1–16.

② Lesley Milroy, *Language and Social Networks*, Oxford: Blackwell, 1987.

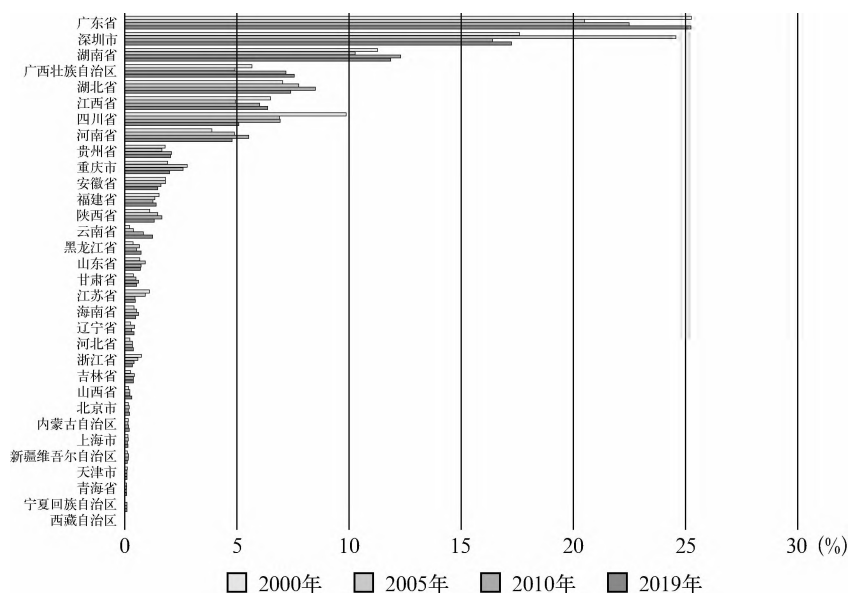


图1 深圳常住人口的户籍登记地(2000—2019)

数据来源:2000年深圳市第五次全国人口普查数据;2005年深圳市全国1%人口抽样调查数据;2010年深圳市第六次全国人口普查数据;2019年深圳市网格办人口统计数据。

以往关于移民聚集的研究大多忽视了语言文化纽带在社会网络中的作用。整体来说,中国的方言具有两大特点:数量多与差异大。在数量方面,根据中国社会科学院及澳大利亚人文学院的报告,中国的七大方言区共有105种细分方言。^①对此,有学者提出中国方言的数量之多、承载的文化之重在世界范围内属于罕见。^②从内容来看,中国方言的内容差异不仅体现在省与省之间,更为独特的是,同省内部的方言差异同样巨大。以广东省为例,其内部的潮汕文化、客家文化与雷州文化差异明显^③,如此多元的方言文化,是“族群隔离”研究中不可忽视的重

① Chinese Academy of Social Sciences, Australian Academy of the Humanities, *Language Atlas of China*, Hong Kong: Hong Kong Longman Publishing (Far East) Limited Company, 1987.

② 李诗旻:《广东粤语的发展及岭南文化的传承和保护》,《学术评论》2012年第3期。

③ 叔翼健:《媒体文化与城市新移民地域文化认同的建构——以广州为例》,《新闻研究导刊》2018年第16期。

要维度。

图 2 汇总了深圳市各类方言使用者的人口比例,由图可知,除普通话之外,客家、粤、赣、潮州、湘、闽、吴、晋等方言均是深圳市方言的重要组成部分。其中,客家、粤、赣、潮州、湘和闽方言的使用人数占比较高,分别为 8.4%、8%、7.4%、7.2%、6.4%及 3.2%,为深圳市的六种主要方言。

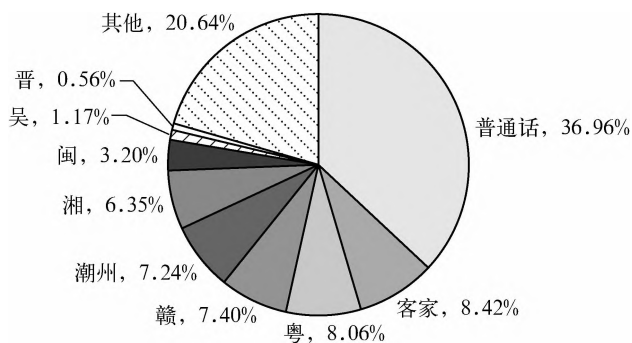


图 2 2005 年深圳各类方言人口比例
数据来源:2005 年深圳市全国 1% 人口抽样调查数据。

本文重点关注客家、粤、赣、潮州、湘和闽六大方言群体的空间聚集、族群经济与劳动收入。本文的具体研究问题为:一、深圳市不同方言群体的空间聚集、族群经济与劳动收入的关系有何异同;二、空间同化理论和分层同化理论对这些群体是否适用。

四、数据、变量和分析方法

本研究采用深圳市 2005 年全国 1% 人口抽样调查数据,该数据总共收集了 193 587 位受访者的信息。按来源地统计,受访者共计来自全国 31 个省级行政区、2509 个县级行政区,共同构成深圳市 2005 年总人口的 2.5%。按户籍人口与流动人口的比例统计,在 193 587 位受访者中,流动人口占绝大多数,共计 80.4%,这些受访者的来源地几乎覆盖

全国绝大部分地区。该数据采集了个人层面的多类信息,包括年龄、性别、户籍地、受教育程度、职业、收入、在深居住时长、常住地等。

本研究的关注点是方言群体的居住聚集和行业聚集与收入之间的关系。这里涉及两个重要的界定:第一,在社区和行业层面,界定某个社区/行业是某个方言群体聚集的社区/行业;第二,在个人层面,界定最终分析样本里的个人是否在自己所属方言群体的聚集区里居住,以及是否在自己所属方言群体的聚集行业里工作。

关于第一个界定,我们使用了全样本数据(即193 587位受访者,包括所有流动人口以及具有深圳户口的居民)。对每一个受访者,我们根据户籍地信息与《中国语言地图集》进行方言匹配,确定其所属方言群体。^①然后,我们以社区和行业为分析单位,运用比值比的方法分别对客家、粤、赣、潮州、湘和闽六大方言群体界定聚集社区和聚集行业。

比值比方法主要分为两个步骤:第一步为计算比值比,其结果反映的是相比于其他社区/行业,某个方言群体在特定社区/行业的相对分布情况;第二步为设立临界值,超过临界值的社区/行业被界定为聚集社区/聚集行业。^②在本文中,计算居住聚集时以居委会为分析单位,一共包括325个居委会;计算行业聚集时参照中国标准行业分类系统中的两位数行业类别,一共包括94个行业(这部分使用了全样本中有工作的人口进行计算)。

比值比的计算公式如下:

$$\text{比值比} = \frac{\frac{E_{ij}}{O_{ij}}}{\frac{E_{\cdot j}}{O_{\cdot j}}}$$

① 《中国语言地图集》第2版,《方言》2014年第1期。

② Richard Alba, John Logan, Kyle Crowder, "White Ethnic Neighborhoods and Assimilation: The Greater New York Region, 1980 - 1990", *Social Forces*, Vol. 75, No. 3, 1997, pp. 883 - 912; John Logan, Richard Alba, Thomas McNulty, "Ethnic Economies in Metropolitan Regions: Miami and Beyond", *Social Forces*, Vol. 72, No. 3, 1994, pp. 691 - 724; Xie Yu, Margaret Gough, "Ethnic Enclaves and the Earnings of Immigrants", *Demography*, Vol. 48, No. 4, 2011, pp. 1293 - 1315.

其中, i 代表方言群体, 分别为客家、粤、赣、潮州、湘和闽; j 表示居委会代码(1, 2, 3, …, 325)或者行业代码(1, 2, 3, …, 94)。分子(E_{ij} 与 O_{ij} 的比值)表示 i 方言在 j 居委会/行业中的代表程度, 等于 j 居委会/行业中 i 方言的人数除以所有其他(除 i 方言以外)人数; 分母(E_{ij}^- 与 O_{ij}^- 的比值)表示 i 方言在所有其他居委会/行业(除去 j 居委会或者 j 行业)中的代表程度, 等于在所有除 j 以外的居委会/行业中 i 方言的人数与其他所有人数的比值。简单来说, 比值比反映的是某个群体在某社区/行业的相对分布情况(相对于在其他社区/行业的分布)。

以居住聚集的计算为例。对于客家人(i =客家人), 我们对第一个居委会($j=1$)计算客家人数与非客家人数的比值(即上面公式的分子 E_{ij} / O_{ij} , 第一个比值), 然后计算在所有其他 324 个社区中, 客家人数与非客家人数的比值(即公式的分母 E_{ij}^- / O_{ij}^- , 第二个比值); 然后把这两个比值相除, 得到客家人在第一个居委会的比值比。同样的方法, 对客家人计算在第二个居委会的比值比, 一直计算到客家人在第 325 个居委会的比值比。以此类推, 对粤、赣、潮州、湘和闽等其他五大方言群体也如此计算。行业聚集也是通过同样的方式计算, 唯一区别在于我们将样本分为了两部分: 雇主(包括自雇)和雇员。

按上述方法, 得到了一系列比值比之后, 下一步就是设置临界值以确定哪些社区/行业是某个方言群体(客家、粤、赣、潮州、湘或闽)所聚集的社区/行业。依据之前的研究, 我们把社区聚集的临界值设为 2, 把行业聚集的临界值设为 1.5。^① 举例, 当某个居委会对客家人来说其比值比大于或等于 2 时, 它就被界定为客家人的聚集社区。需特别说明的是, 为了使行业聚集的界定更为严谨, 只有雇主样本的比值比和雇员样本的比值比都大于或等于 1.5 时, 该行业才会被界定为对应群体(比如客家人)聚集的行业。

① Per-Anders Edin, Peter Fredriksson, Olof Åslund, "Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment", *The Quarterly Journal of Economics*, Vol. 118, No. 1, 2003, pp. 329 - 357; John Logan, Richard Alba, Thomas McNulty, "Ethnic Economies in Metropolitan Regions: Miami and Beyond", *Social Forces*, Vol. 72, No. 3, 1994, pp. 691 - 724.

针对各方言群体的聚集社区和聚集行业确定之后,第二个界定是针对样本里面的流动人口,确定他/她是否在自己所属方言群体的聚集区里居住,以及是否在自己所属方言群体的聚集行业里工作。在这一步,我们把分析限定在属于客家、粤、赣、潮州、湘或闽等六大方言群体的流动人口之中。“居住聚集”和“行业聚集”这两个核心自变量都是因变量。如果受访者在所属方言群体的聚集社区居住,那么“居住聚集”这个变量就等于1,否则等于0;如果受访者在所属方言群体聚集的行业工作,那么“行业聚集”这个变量就等于1,否则等于0。

本文的因变量是受访者月收入的对数。除了“居住聚集”和“行业聚集”之外,我们还控制了年龄(及其平方)、性别、婚姻状况、教育水平、在深居住时长、雇佣身份、居住社区的社会经济指数、职业大类和行业大类等变量。所有变量及其取值见表1。

我们把最终的分析样本限定在六大方言群体(客家、粤、赣、潮州、湘和闽)中有收入且年龄在16—60岁的受访者,一共57 568人(其中客家11 719人、粤11 916人、赣11 555人、潮州7 697人、湘10 007人、闽4 674人)。在多变量分析中,我们使用了最小二乘法回归模型(OLS regression)。我们对总的分析样本以及每个方言群体的分样本做多变量分析,以比较六大方言群体之间的差异。

表1 2005年深圳流动人口六大方言群体居住聚集和行业聚集分析变量表

变量	变量取值
因变量	
月收入的对数	大于0的常数
核心自变量	
居住聚集	1:在所属方言群体聚集区居住 0:在非聚集区居住
行业聚集	1:在所属方言群体聚集行业工作 0:在非聚集行业工作

续表

变量	变量取值
因变量	
其他自变量	
年龄	大于 0 的常数
年龄平方	大于 0 的常数
性别	1：女性；0：男性
婚姻状况	1：已婚；0：未婚
高等教育	1：已接受高等教育 0：未接受高等教育
在深居住时长	1：小于一年 0：大于等于一年
雇佣身份	1：雇员 2：雇主 3：自雇
居住社区的社会经济指数	0 到 100 之间的常数
职业类别	1：管理人员 2：专业人员 3：文员 4：商业/服务业人员 5：体力劳动者
行业类别	1：制造业 2：建筑业 3：批发/零售 4：住宿/餐饮 5：其他

五、分析结果

（一）深圳六大方言群体所聚集的社区与行业

以往研究揭示深圳市存在数以百计的城中村，而城中村很多情况下也是“同乡村”。^① 我们对深圳流动人口六大方言群体居住聚集的分析也反映

^① 周春山、杨高、王少剑：《深圳农民工集聚空间的演变特征及影响机制》，《地理科学》2016 年第 11 期。

了这一点。深圳很多社区存在一个或多个方言群体聚集居住的情况。从行政区域来看,福田区和罗湖区内的方言聚集区呈散点分布,与之相比,处于外围的宝安、光明、龙华和龙岗等聚集社区连片分布,且数量更多。从各居委会内方言群体聚集的数量来看,处于中心的福田、罗湖和南山等具有较少数量的多群体聚集社区,而外围的宝安、光明、龙华和龙岗等具有较多数量的多群体聚集社区。这可能跟行政设置、产业布局和住房供应等相关。福田区和罗湖区为深圳市的行政中心和商业中心;而外围区域则有更多的工业区,比如2005年宝安区和龙岗区的工业园区数量占全市的86%,此外住房成本较低也是流动人口聚集于宝安和龙岗等区域的重要原因。

整体来看,客家、粤、赣、潮州、湘和闽等六大方言群体的空间聚集呈现出不同的分布规律。比如,客家人主要聚集在龙岗和宝安,粤群体主要聚集在宝安、光明和龙华,闽群体主要聚集在龙岗和坪山;而除了外围的龙岗和大鹏新区,潮州人在市中心的福田和罗湖也有不少聚集社区。这些不同方言群体居住聚集的空间分布差异,说明对我国流动人口的居住聚集及其相关结果进行研究时,需重视语言文化群体的界定和比较。

表2列出了各方言群体所聚集的行业。从聚集行业的数量来看,客家、粤、赣群体分别有4个聚集行业,潮州群体有3个,闽群体有2个,湘群体有1个。从聚集行业的种类来看,赣群体聚集于城市公共交通业,居民服务业,农副食品加工业,木材加工及木、竹、藤、棕、草制品业;客家群体聚集于有色金属冶炼及压延加工业、零售业、娱乐业及邮政业;粤群体聚集于渔业、农业、畜牧业及建筑装饰业;潮州群体聚集于农副食品加工业、零售业及非金属矿物制品业;闽群体聚集于家具制造业及餐饮业;湘群体聚集于农业。

表3汇总了各方言群体居住聚集与行业聚集的整体情况。居住聚集的数据显示,潮州和客家群体的聚集社区数量最多,均为40个,所对应的人数占比分别为35.3%和27%;其次为粤和闽群体,分别为34个和32个聚集社区,对应的聚集人口比例分别为23.1%和22.9%;湘和赣群体的聚集社区数量分别为25个和20个,对应的聚集人口比例分别

为 14.4%和 16.2%。行业聚集的情况也反映出流动人口内部的异质性，潮州和客家群体依然是六大方言群体中聚集趋势最明显的，分别有 42.8%和 19.4%的人在所属方言群体聚集的行业工作，远远高于其他方言群体（闽 10.8%、赣 8.7%、粤 4.3%、湘 0.1%）。

表 2 2005 年深圳流动人口六大方言群体聚集的行业

方言群体	该群体所聚集的行业
客家	有色金属冶炼及压延加工业、零售业、娱乐业、邮政业
粤	渔业、农业、畜牧业、建筑装饰业
赣	城市公共交通运输业，居民服务业，农副食品加工业，木材加工及木、竹、藤、棕、草制品业
潮州	农副食品加工业、零售业、非金属矿物制品业
湘	农业
闽	家具制造业、餐饮业

表 3 2005 年深圳流动人口六大方言群体的居住聚集与行业聚集情况

方言群体	居住社区总数	聚集社区总数	居住于方言聚集社区的人口比例	所在行业总数	聚集行业总数	就业于方言聚集行业的人口比例
客家	323	40	27.0%	84	4	19.4%
粤	324	34	23.1%	85	4	4.3%
赣	323	20	16.2%	77	4	8.7%
潮州	317	40	35.3%	77	3	42.8%
湘	324	25	14.4%	80	1	0.1%
闽	317	32	22.9%	76	2	10.8%

（二）深圳六大方言群体居住聚集和行业聚集与收入的关系

表 4 列出了最终分析样本的总体特征，以及按是否在聚集社区居住和是否在聚集行业工作进行分类的情况。在居住聚集方面，住在所属方言群体聚集区

内的人,相比住在非聚集区的人,平均月收入稍微高些(分别为 1553 元和1389.7 元)。在行业聚集方面,在所属群体聚集行业工作的人,相比在非聚集行业工作的人,平均收入却更低(分别为 1483.6 元与 1743.3 元)。

表 4 2005 年深圳流动人口六大方言群体的描述性统计

	六大方言群体全部样本	在聚集社区居住		在聚集行业工作	
		是	否	是	否
月收入	1516.3 (1772.3)	1553.0 (1671.5)	1389.7 (2077.3)	1483.6 (1754.8)	1743.3 (1873.6)
年龄	27.4 (8.3)	27.4 (8.2)	27.3 (8.7)	27.0 (8.1)	29.7 (9.3)
女性	44.5	44.1	46.0	45.0	40.6
已婚	42.7	42.6	42.8	40.9	55.0
高等教育	9.8	11.5	4.0	10.7	3.5
在深居住少于 1 年	24.9	25.2	24.0	25.6	20.6
雇佣身份					
雇员	87.6	89.0	84.0	91.3	61.7
雇主	4.7	4.3	6.1	3.5	13.4
自雇	7.7	7.1	9.9	5.3	24.9
职业类别					
管理人员	3.5	4.0	2.2	3.3	4.3
专业人员	6.3	7.0	3.1	7.0	1.4
文员	7.4	8.0	5.0	8.0	2.6
商业/服务业人员	34.4	33.9	35.9	28.1	77.8
体力劳动者	48.6	47.1	53.8	53.6	14.0
行业类别					
制造业	49.9	49.0	53.0	56.4	4.8
建筑业	3.7	4.0	2.9	3.5	5.2
批发/零售	19.5	18.5	22.9	12.4	68.5
住宿/餐饮	6.7	7.1	5.4	6.9	5.4
其他	20.2	21.4	15.9	20.8	16.1

续表

	六大方言群体全部样本	在聚集社区居住		在聚集行业工作	
		是	否	是	否
方言群体					
客家	20.4	19.3	23.9	19.0	30.1
粤	20.7	20.3	22.2	22.9	5.8
赣	20.1	22.1	13.2	21.0	13.9
潮州	13.4	11.4	20.1	9.1	43.4
湘	17.4	19.0	11.8	19.9	0.1
闽	8.1	7.9	8.8	8.3	6.8
样本量	57 568	44 612	12 956	50 320	7248

注：收入和年龄的数字为均值(标准差)，其他变量的数字为百分比。

对六大方言群体分别进行分析发现，居住聚集与收入之间的关系存在差别。对于客家、粤、湘和闽群体来说，居住在所属群体聚集区内的人，相比住在非聚集区的，平均收入更低(客家：1328.1 元与 1497.4 元；粤：1022.6 元与 1556.1 元；湘：1089.2 元与 1485.2 元；闽：1282.2 元与 1709.1 元)；但是对赣和潮州群体来说，住在聚集区内的平均收入却比在非聚集区的更高(赣：1596.9 元与 1415.4 元；潮州：1954.8 元与 1912.6 元)。行业聚集与收入之间的关系在六大方言群体间也有差别(客家：按是否在聚集行业工作的平均月收入为 1591.3 元和 1420.9 元；粤：1695.2 元与 1417.8 元；赣：1518.7 元与 1434.9 元；潮州：2011.5 元与 1868.6 元；湘：1777.8 元与 1424.6 元；闽：1203 元与 1651.6 元)。

表 5 展示了 2005 年深圳六大方言流动人口的收入回归 (OLS regression) 分析结果。模型 1 是全样本分析，模型 2—7 分别为客家、粤、赣、潮州、湘和闽的分样本分析。如表所示，在控制了其他变量之后，居住聚集和行业聚集与收入之间的关系在各方言群体之间存在差异。对客家、粤和闽来说，在所属方言群体聚集区(相比在非聚集区)居住的人平均收入更低(分别低 3.5% [$1 - e^{-0.036}$]、8.9% [$1 - e^{-0.093}$] 和 7.3% [$1 - e^{-0.076}$])；而赣和潮州则相反，在所属方言群体聚集区居住的平均收入更高(分别高 12.6%

表 5 2005 年深圳流动人口人口六大方言群体的收入回归分析结果

	模型 1	模型 2 (客家)	模型 3 (粤)	模型 4 (赣)	模型 5 (潮州)	模型 6 (湘)	模型 7 (闽)
居住聚集	-0.008 (0.005)	-0.036*** (0.010)	-0.093*** (0.010)	0.119*** (0.012)	0.034* (0.014)	-0.001 (0.012)	-0.076*** (0.017)
行业聚集	-0.011 (0.007)	-0.098*** (0.020)	0.106** (0.032)	-0.011 (0.018)	-0.144*** (0.020)	0.068 (0.145)	0.005 (0.039)
年龄	0.051*** (0.002)	0.062*** (0.004)	0.058*** (0.003)	0.041*** (0.004)	0.059*** (0.005)	0.039*** (0.004)	0.083*** (0.006)
年龄二次项	-0.076*** (0.002)	-0.092*** (0.006)	-0.084*** (0.005)	-0.068*** (0.006)	-0.083*** (0.007)	-0.061*** (0.006)	-0.116*** (0.009)
女性	-0.161*** (0.004)	-0.180*** (0.009)	-0.170*** (0.009)	-0.137*** (0.009)	-0.185*** (0.014)	-0.101*** (0.009)	-0.153*** (0.015)
已婚	0.055*** (0.006)	0.031* (0.014)	0.081*** (0.013)	0.069*** (0.013)	0.067** (0.021)	0.028* (0.014)	0.026 (0.023)
高等教育	0.463*** (0.008)	0.444*** (0.018)	0.481*** (0.017)	0.497*** (0.015)	0.361*** (0.030)	0.482*** (0.016)	0.457*** (0.026)
在深居住少于 1 年	-0.159*** (0.005)	-0.136*** (0.011)	-0.126*** (0.010)	-0.178*** (0.010)	-0.184*** (0.016)	-0.161*** (0.010)	-0.151*** (0.016)
雇佣身份(参照组:雇员)							
雇主	0.419*** (0.010)	0.399*** (0.022)	0.378*** (0.025)	0.328*** (0.027)	0.453*** (0.022)	0.375*** (0.032)	0.488*** (0.035)
自雇	0.250*** (0.008)	0.216*** (0.018)	0.197*** (0.019)	0.094*** (0.020)	0.369*** (0.019)	0.161*** (0.024)	0.253*** (0.031)
社区社会经济指数	0.007*** (0.000)	0.007*** (0.000)	0.006*** (0.000)	0.007*** (0.000)	0.006*** (0.000)	0.007*** (0.000)	0.006*** (0.000)

续表

	模型 1	模型 2 (客家)	模型 3 (粤)	模型 4 (赣)	模型 5 (潮州)	模型 6 (湘)	模型 7 (闽)
职业类别(参照组:管理人员)							
专业人员	-0.284*** (0.014)	-0.312*** (0.032)	-0.316*** (0.030)	-0.324*** (0.030)	-0.219*** (0.042)	-0.290*** (0.033)	-0.165*** (0.045)
文员	-0.449*** (0.013)	-0.416*** (0.031)	-0.475*** (0.029)	-0.484*** (0.030)	-0.416*** (0.039)	-0.512*** (0.032)	-0.363*** (0.044)
商业/服务业人员	-0.609*** (0.012)	-0.557*** (0.028)	-0.688*** (0.026)	-0.686*** (0.027)	-0.517*** (0.030)	-0.633*** (0.030)	-0.500*** (0.037)
体力劳动者	-0.570*** (0.012)	-0.540*** (0.029)	-0.657*** (0.027)	-0.526*** (0.028)	-0.588*** (0.034)	-0.595*** (0.031)	-0.547*** (0.039)
行业类别(参照组:制造业)							
建筑业	0.049*** (0.011)	0.015 (0.030)	-0.010 (0.027)	0.049* (0.025)	0.045 (0.031)	0.020 (0.027)	0.023 (0.042)
批发/零售	0.135*** (0.008)	0.130*** (0.023)	0.148*** (0.017)	0.227*** (0.017)	0.081** (0.027)	0.116*** (0.019)	0.094*** (0.027)
住宿/餐饮	-0.031** (0.010)	-0.047* (0.020)	-0.022 (0.020)	0.055* (0.023)	-0.168*** (0.033)	-0.056* (0.023)	-0.071 (0.045)
其他	0.083*** (0.006)	0.036** (0.014)	0.076*** (0.014)	0.207*** (0.014)	-0.070** (0.024)	0.096*** (0.014)	0.083*** (0.023)
常数项	6.673*** (0.029)	6.521*** (0.067)	6.631*** (0.061)	6.833*** (0.063)	6.588*** (0.087)	6.905*** (0.068)	6.120*** (0.102)
样本量	57 568	11 719	11 916	11 555	7697	10 007	4674
调整 R ²	0.427	0.372	0.486	0.436	0.390	0.439	0.498

注:括号内为标准误;* p<0.05,** p<0.01,*** p<0.001。

$[e^{0.119} - 1]$ 和 $3.5\%[e^{0.034} - 1]$);对于湘群体来说,居住聚集与收入之间的关系在统计上不显著($\alpha=0.1$)。在行业聚集方面,就粤群体而言,在所属方言群体聚集行业(相比在非聚集行业)工作的人平均收入更高(高 $11.2\%[e^{0.106} - 1]$);而客家和潮州则相反,在聚集行业工作的成员平均收入更低(分别低 $9.3\%[1 - e^{-0.098}]$ 和 $13.4\%[1 - e^{-0.144}]$)。

为了更直观地比较不同方言群体之间居住/行业聚集与收入关系的差异,以及对空间同化理论和分层同化理论的回应情况,我们把表5的回归分析结果概括为表6。如表所示,对各方言群体分别进行的实证分析结果,说明空间同化理论对客家、粤和闽适用,对这些方言群体而言,聚集的社区起到了过渡平台的作用;从移民融入的角度看,这些聚集的社区很可能为新来者提供了“庇护所”,但当这些成员获得更高的收入之后,则逐渐迁出聚集区,融入主流或条件更好的社区(在截面数据分析结果中,我们对应地发现在聚集社区居住的平均收入更低,在非聚集区居住的收入更高)。然而,对赣和潮州群体来说,分析结果与空间同化理论并不一致。对分层同化理论的回应方面,粤群体证实了族群经济与收入的正相关,而客家和潮州群体则呈现出负相关。这些比较分析结果再次说明,在研究流动人口的族群聚集及相关议题时,需要重视流动人口内部的异质性。

表6 2005年深圳流动人口方言群体聚集与收入关系比较

方言群体	居住聚集	空间同化理论	行业聚集	分层同化理论
客家	—	过渡社区	—	负相关
粤	—	过渡社区	+	正相关
赣	+	不一致	/	/
潮州	+	不一致	—	负相关
湘	/	/	/	/
闽	—	过渡社区	/	/

注:“+”表示族群聚集与收入正相关;“—”表示族群聚集与收入负相关;“/”表示族群聚集与收入之间的关系在统计学上不显著($\alpha=0.1$)。

六、总结与讨论

居住空间聚集和族群经济是移民研究的重点。以往关于中国城市族群聚集的研究大都把流动人口当成一个整体,忽略了移民来源地广泛及方言文化多样带来的内部差异。在这个背景下,这篇文章强调“方言文化差异”的视角,利用深圳市 2005 年全国 1% 人口抽样调查数据,比较了客家、粤、赣、潮州、湘和闽等六大方言群体的居住聚集和行业聚集情况,以及两者与收入之间的关系,进而检验空间同化理论和分层同化理论对中国城市不同方言群体的适用性。

我们首次使用比值比的方法对上述六大方言群体分别界定其所聚集的社区和行业,发现这些群体的空间聚集形态和行业聚集情况各有差异。接着我们对这六大群体分别进行收入回归分析,发现各方言群体的居住聚集、行业聚集与收入的关系也存在差异。

对客家、粤和闽群体而言,居住聚集和收入负相关,那些在所属群体聚集的社区居住的人,相比在非聚集区居住者,平均月收入更低。这个结果与空间同化理论一致。对客家、粤和闽这些群体而言,其方言聚集社区类似于“过渡社区”,可能为新来者提供“庇护”,随着成员社会经济地位的提高,他们会迁出这些聚集地进而融入主流或条件更好的社区。但是,对赣和潮州群体而言,居住聚集和收入正相关,那些在所属群体聚集的社区(相比在非聚集区)居住的人,平均月收入更高。这些结果说明,空间同化理论的适用性因方言文化群体而异。

在行业聚集与收入的关系方面,我们的实证分析跟以往分层同化理论的文献一样,发现同时存在正相关和负相关的情况。对粤群体而言,在所属方言群体聚集行业(比在非聚集行业)工作的人,平均收入更高;而客家和潮州群体则相反,在聚集行业工作的人平均收入更低。这些结果进一步说明,中国城市移民研究需要关注方言文化差异。

本文对中国城市族群聚集和移民融入等相关研究的贡献,在于提出

了“方言文化差异”的视角,并以深圳为例,说明经典的同化理论在中国城市的适用性并不能一概而论,需考察流动人口的内部差异。基于截面数据,这篇文章描述的是空间聚集、行业聚集与收入之间的相关关系。族群聚集与收入之间的因果效应,需要更合适的研究设计和量化数据进行探讨,这是本文的局限,也是我们未来研究努力的方向。

为何我国城市流动人口不同方言群体之间会存在上述差异?由于数据限制,这篇文章无法进行实证分析。我们认为,大概可以从文化、行业与社会空间分析的角度促进相关研究的发展。

从文化角度来看,弱连接理论(strength of weak ties)和结构洞理论(structural hole theory)可能提供一些参考。这两个理论指出过于紧密的社会网络反而可能不利于组织的绩效和发展。^①以往的相关研究发现“家长式”运作关系存在于紧密的族群网络和族裔经济中,移民企业家可能压迫自己族群的工人,给他们工作时间长、工资相对较低的职位。^②此外,根据文献记载,粤文化相对开放,而潮州文化和客家文化则相对保守,浓厚的乡族观念使得客家和潮州群体的内部关系十分紧密^③,是否过于紧密的宗族网络在一定程度上阻碍了“庇护所”的发展,变“庇护”为“剥削”,这有待考察。

① James Montgomery, “Job Search and Network Composition: Implications of the Strength-Of-Weak-Ties Hypothesis”, *American Sociological Review*, Vol. 57, No. 5, 1992, pp. 586-596; David Constant, Lee Sproull, Sara Kiesler, “The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice”, *Organization Science*, Vol. 7, No. 2, 1996, pp. 119-135; Ronald Burt, “Structural Holes and Good Ideas”, *The American Journal of Sociology*, Vol. 110, No. 2, 2004, pp. 349-399.

② Edna Bonacich, “‘Making It’ in America: A Social Evaluation of the Ethics of Immigrant Entrepreneurship”, *Sociological Perspectives*, Vol. 30, No. 4, 1987, pp. 446-466; Jimmy Sanders, Victor Nee, “Limits of Ethnic Solidarity in the Enclave Economy”, *American Sociological Review*, Vol. 52, No. 6, 1987, pp. 745-773; Eric Fong, Feng Hou, “Effects of Ethnic Enclosure of Neighborhoods, Workplace, and Industrial Sectors on Earnings”, *Social Science Research*, Vol. 42, No. 4, 2013, pp. 1061-1076.

③ 李新梅:《从粤语的谐音特点看广州文化》,《深圳职业技术学院学报》2016年第4期;李滢锋、古亮锋、庞颖怡:《近代以来梅州客家文化发展的传承与发展研究》,《今古文创》2021年第31期;胡少东、李龙、黄元盈:《乡族关系、商人信仰与近代潮汕商帮治理》,《南大商学评论》2020年第4期。

行业特征可能也是方言群体之间差异的来源。以潮州群体为例,根据本研究的统计,潮州群体中接近一半的人数(42.8%)就业于方言聚集行业,即农副食品加工业、零售业和非金属矿物制品业,且其中约40%的人员为雇主(含自雇),这两个比例均远远高于其他方言群体。除此以外,通过对比各群体的人口分布可知,其他方言群体(如客家、粤、湘)主要居住于以工业区为主的区域,但潮州群体的聚集社区也分布于中心区域(福田和罗湖)。行业性质及较高的雇主占比是否导致行业聚集与收入之间的负相关,值得探究。

从社会空间分析的角度看,族群聚集是多层次、多维度共演的动态过程。宏观层面的规划设计和产业分布等,与微观层面各群体的社会网络、机会成本和利弊权衡等,形成社会空间大网络,共同对流动人口的居住和行业聚集起作用。中国城市研究的一大挑战是打破学科壁垒。我们认为,未来城市实证研究应重视跨学科的“社会空间分析”视角,结合人口调查数据与外部空间数据,更全面和深入地研究族群空间分布与其他经济社会网络的形成、发展及影响。

Birds of a Feather: Chinese Dialects, Residential and Industrial Enclaves, and Migrants' Earnings in Shenzhen, China

Pu Juanbo Zeng Donglin Zhang Zhuoni

Abstract: Residential segregation and ethnic enclaves have received much attention in migration and urban studies. Internal migrants in urban China are often classified into dialect clusters given great varieties of places of origin and related regional differences in language and culture. Yet few research has empirically compared migrants of different dialects regarding residential and industrial enclaves and their relationships with labor market outcomes. In this paper, we filled the gap by using data from the 2005 Shenzhen Mini Census to compare among Hakka, Cantonese, Gan, Chiu Chow, Xiang and Min migrants, to test whether the spatial assimilation theory and segmented assimilation theory are applicable to these different dialect groups. We found that for Hakka, Cantonese and Min dialect groups, migrants who lived in residential enclaves earned less than those who did not, yet for Gan and Chiu Chow the corresponding relationship were positive. Furthermore, consistent with previous literature, we found that industrial enclaves were positively associated with earnings for some dialect group (i. e. , Cantonese) and negatively associated with earnings for some others (i. e. , Hakka and Chiu Chow). This paper contributes to the literature with a new “dialect comparison” perspective to examine the internal variation of migrants in urban China. Limitations of this study and possible future research topics were also discussed.

Keywords: residential enclave, industrial enclave, earnings, dialect comparison perspective, spatial assimilation theory, segmented assimilation theory