

贵州农贸市场信息采集人员 信息上报行为的大数据分析研究

周治黔 吴海 令狐昌平

(贵州省农村综合经济信息中心 贵州贵阳 550081)

【摘要】：大数据时代的到来，给研究特定人群的工作或生活行为带来了契机。随着贵州省农村综合经济信息中心收集农贸市场信息采集人员上报的价格信息数据量的不断增加，为研究此类特定人群的行为规律奠定了基础。运用大数据统计分析方法，对两位信息采集人员信息上报记录的年总次数、月次数、周分布、节假日分布以及时间点进行分析，得出八寨镇信息员的信息上报行为比较有规律性，而虾子镇信息员的上报行为随意性较大；在节假日和周末，两名信息员都有度假和休息的习惯以及虾子镇信息员每天的工作时间要普遍早于八寨镇信息员，下班时间较晚的情况等行为规律。并基于上述分析结论，建议管理者需进一步加强对信息采集人员信息上报行为的规范化管理，强化节假日信息采集上报工作。

【关键词】：信息上报；行为规律；大数据；分析研究

【中图分类号】：F323.7 **【文献标识码】**：A

大数据时代的到来，是信息技术快速发展的一个重要时期，给研究特定人群的工作或生活行为带来了契机。孟祥玉等人利用微博数据开展了北京地区居民时空行为的大数据分析，总结了热点分布区域和不同时段微博热点分析，唐朝军等人利用微博数据开展了高校学生微博使用行为大数据分析和研究，总结了高校学生微博使用行为规律。随着贵州省农村综合经济信息中心（以下简称农经中心）收集整理的农贸市场信息采集人员信息上报次数的不断增加，为此类负责信息采集的特定人群的行为规律研究奠定了数据基础。本文将利用贵州省农贸市场信息采集人员信息上报记录的具体时间，应用统计分析方法，通过分析其中两位信息采集人员（以下简称信息员）信息上报记录的时间序列，开展行为痕迹分析，尝试了解和掌握信息员工作的时间规律，以利于有针对性地提升信息采集工作的规范化管理水平，进一步增强信息员所采集数据的时空代表性。

1 资料来源

本文选取了农经中心收集整理的毕节市八寨镇牲畜交易市场信息采集人员（以下简称八寨镇信息员）和遵义市遵义县虾子镇中国辣椒城信息采集人员（以下简称虾子镇信息员）的信息上报记录的时间序列作为本次分析研究的对象。两组信息上报记录的时间跨度为2011年1月至2013年12月，记录的总次数分别为692次和560次。

收稿日期：2018-12-26

基金项目：贵州省科技计划课题（课题名称：贵州省农村信息化工程技术研究中心，课题编号：黔科合农G字[2014]4001）。

作者简介：周治黔（1979—），男，贵州贵阳人，工程师，研究方向：农村综合经济信息。

2 数据预处理与分析

2.1 研究对象行为一致性分析

鉴于需对上述两名信息员的信息上报行为进行分析，须确保行为人为人的一致性。根据农经中心信息管理部门的反馈，上述两名信息员均是自 2010 年以来一直从事所负责的农贸市场的信息采集上报工作，分析对象的上报行为在所研究时段内是保持一致的。

2.2 时间序列完整性分析

考虑到本次对比分析是依据信息上报记录的时间序列来进行的，因此，如果某一时间序列在月度上出现缺失，则认定其时间序列不完整，影响分析结果，不能用于本次分析。根据对两组信息上报记录的时间序列的分析，两组时间序列在 2011 年 1 月至 2013 年 12 月期间每月均有数据，时间序列基本完整，可以作为本次对比分析的研究对象。

2.3 时间序列可靠性分析

对拟分析信息上报记录的时间序列中的时间数据进行逐一甄别，每个时间数据均为两名信息员上报信息到农经中心服务器的时间，均以服务器的同一个时钟作为记录时间的依据，具有一致性，且两个时间序列中的年、月、日、时、分记录清晰，数据可靠。

2.4 时间序列合理性分析

在两组信息上报记录的时间序列中，每个时间数据均以日为时间单元，即每日仅有一条时间数据，不存在一日多报或一日多条不同时间记录的现象。通过数据清洗和甄别，两组时间序列中均未出现同一日或同一条记录重复登记的现象，可以作为研究对象进行分析。

3 数据对比分析结果

3.1 年次数对比分析

两组信息上报记录的时间序列显示，在分析时段（3 个完整年度）内的记录总次数分别为 692 次和 560 次，年度总次数平均分别为 230.67 次和 186.67 次，差异较大。其中虾子镇信息员 2013 年信息上报量较少，仅为 105 次（详见表 1）。分析结果表明，虾子镇信息员信息上报工作的随意性要比八寨镇信息员要大一些。

表 1 2011 年 1 月至 2013 年两名信息员年度信息上报次数统

计表		
年总数	八寨镇信息员（次）	虾子镇信息员（次）
2013 年	244	105
2012 年	214	212
2011 年	234	243
合计	692	560
平均值	230.67	186.67
标准差	15.28	72.40

3.2 月次数对比分析

3.2.1 月均态分析。对两组信息上报记录的时间序列按月（共 36 个月）进行分析，每月平均信息上报次数分别为八寨镇信息员 9~23 次，辣椒城信息员 2~23 次；平均值分别为八寨镇信息员 19.22 次，虾子镇信息员 15.56 次；标准差分别为八寨镇信息员 3.02 次、虾子镇信息员 6.26 次。相比之下，虾子镇信息员信息上报次数离散度较高（详见图 1）。

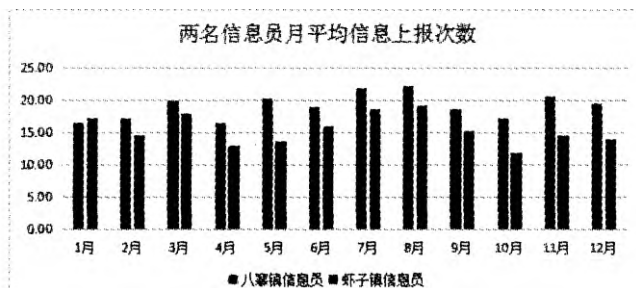


图 1 2011~2013 年两名信息员月平均信息上报次数

3.2.2 重点时段分析。在此基础上，重点对虾子镇信息员信息上报量偏少的 2013 年进行逐月分析，发现该年度该信息员月信息上报量在 2~20 次之间，月平均信息上报量为 8.75 次，标准差为 5.93 次。该年度其月信息上报量离散度较高。

分析结果表明，八寨镇信息员每月信息上报工作相对比较规范，而虾子镇信息员在规范性方面相对较差。

3.2.3 具体月份及节假日分析。从具体月份的来看，八寨镇信息员月上报量偏少的前四个月份是 1 月、4 月、2 月和 10 月，虾子镇信息员则是 10 月、4 月、5 月和 12 月。

在节假日信息上报工作方面，两名市场信息员在三个年度内的国家法定节假日期间，信息上报都非常少，均只上报了 1 次信息，其中八寨镇信息员上报时间为 2013 年春节，虾子镇信息员上报时间为 2012 年元旦节。两名信息员在清明、五一、端午、中秋和国庆节期间均没有上报过信息。

结合月份和假期所在当月的信息上报量分析，可知：①两名信息员在三个年度的国庆期间月信息上报量均进入了偏少前四位；②八寨镇信息员在三个年度的春节期间月信息上报量均进入了偏少前四位，而虾子镇信息员也有两个年度进入了偏少量前四位；③虾子镇信息员在两个年度的清明节期间月信息上报量进入了偏少前四位，分列第二位和第三位，第三年度月信息上报量也是处于偏少的第五位，而八寨镇信息员

则不明显，仅有一个年度月信息上报量进入偏少前四位，其它两个年度也分别处于第六位和第七位，基本属于正常状态。

分析结果表明，两名信息员均在节假日期间基本都处于休息状态，没有在节假日坚持信息上报的习惯；但虾子镇信息员对清明节的重视程度比八寨镇信息员要高。

3.3 周分布对比分析对两组信息上报记录的时间序列按工作日与非工作日进行分析，两名信息员均明显呈现“工作日上报得多、非工作日上报得少”的特点。其中八寨镇信息员更为显著，星期六上报的信息次数仅为 12 次，星期日上报的信息次数仅为 10 次，周六周日次数占总次数的 3.18%；而虾子镇信息员星期六上报的有 16 次，星期日上报的有 15 次（详见图 2），周六周日次数占总次数的 5.54%。相比之下，虾子镇信息员周末工作的概率略大一些。

分析结果表明，两名信息员对于周末休息应该比较重视。

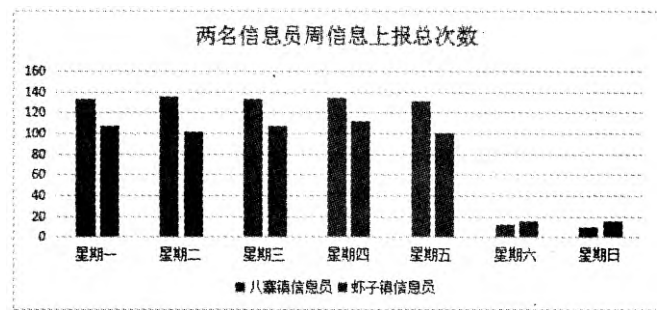


图 2 2011～2013 年两名信息员周信息上报总次数

3.4 小时次数对比分析

对两组信息上报记录的时间序列按信息上报的时间点进行分析，有以下特点。

3.4.1 相同之处。二者在 07 时至 18 时均有信息上报，信息上报的时间普遍集中在上午 08 时至 11 时之间，该时段两名信息员上报信息的数量占总次数的比例分别为：八寨镇信息员 92.49%，虾子镇信息员 88.21%（详见下图 3）。

3.4.2 不同之处。（1）信息上报特别集中时段不一致。八寨镇信息员信息上报的特别集中时段是 09 时至 10 时，而虾子镇信息员则特别集中在 08 时至 09 时。

（2）信息上报最早时间点不一致。八寨镇信息员最早信息上报记录为 06 时 39 分；虾子镇信息员最早信息上报记录为 07 时 08 分。但是八寨镇信息员 08 时以前的信息上报记录仅有 11 次，而虾子镇信息员 08 时以前的记录共有 31 次。

信息上报最晚时间点不一致。八寨镇信息员最晚信息上报时间为 18 时 46 分，而虾子镇市场信息员则在 22 时 03 分，且 18 时 12 分和 19 时 19 分另外还分别各有 1 次信息上报记录。

八寨镇市场信息员信息上报时段呈现双峰现象，上午 09 时至 10 时和下午 14 时；而虾子镇市场信息员

则呈现单峰现象，上午 07 时至 09 时显著增多，随后逐渐减少。

分析结果表明，常态下，虾子镇信息员开展信息上报工作的时间要比八寨镇信息员早，且重点是在上午较早的时间段就完成上报工作；而八寨镇信息员信息上报时段的双峰现象，更加类似于从事按时上下班的人员所为。

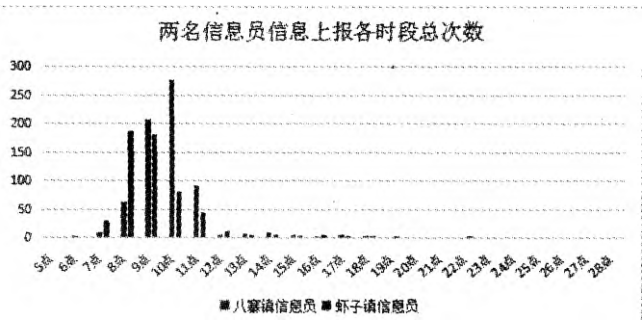


图 3 2011~2013 年两名信息员信息上报各时段总次数

4 结论

通过对两组信息上报记录的时间序列进行对比分析，可以看出，在信息上报行为方面两名信息员既有共同点，也有不同点。

相比较而言，不管是年度信息上报，还是月信息上报，八寨镇信息员的信息上报行为比较有规律性，而虾子镇信息员的信息上报行为随意性较大。

在节假日和周末，两名信息员都有度假和休息的习惯，没有明显的利用节假日和周末进行信息上报的行为；不同的是虾子镇信息员应该对传统的清明节更加重视，该节日相邻时段其未进行信息上报的日数相对较多。

虾子镇信息员每天的工作时间要普遍早于八寨镇信息员，还存在下班时间较晚的情况。而八寨镇信息员的信息上报行为更加具有按时上下班的规律。

5 讨论

本研究仅基于两名信息员三年的上报行为，不一定能够完全表征行为人的行为特点，下一步可扩大研究对象和时间范围进行分析研究，以便更准确地掌握信息员上报行为的规律。在今后的信息上报管理工作中，管理者应进一步强化规范管理，降低行为人的主观随意性给信息数据序列带来难以弥补的缺失。农贸市场的价格波动往往在节假日期间较大，建议管理者应进一步加强在节假日期间的信息采集工作。

参考文献

[1] 孟祥玉, 郑新奇. 北京地区居民时空行为的大数据分析[J]. 测绘学报, 2015, 44 (S0): 1-5.

[2] 唐辉军, 宋扬, 熊松泉. 高校学生微博使用行为大数据分析和研究[J]. 科教文汇, 2015, 322 (A):

122-123.