



پژوهشکده آمار

جمهوری اسلامی ایران
مرکز آمار ایران
پژوهشکده آمار

پیشنهاد طرحی برای برآورد خالص کم‌شماری در سرشماری عمومی نفوس و مسکن سال ۱۳۸۵

تابستان ۱۳۷۹

**پیشنهاد طرحی برای برآورد خالص کم‌شماری
در سرشماری عمومی نفوس و مسکن سال ۱۳۸۵**

محمدرضا فریدروحانی - یدا... واقعی - زهره فلاح محسن‌خانی

گروه پژوهشی طرحهای فنی و روشهای آماری

تابستان ۱۳۷۹

به نام خداوند جان و خرد

پیشگفتار

مبحث برآورد خطای پوشش سرشماری یا همان خالص کم‌شماری از جمله مباحث جدال‌انگیز حاشیه سرشماریهای عمومی نفوس و مسکن بوده است. علیرغم کوششهای فراوان بکار گرفته شده و کسب تجربیات بسیار جهت کاستن از خطاهای پوشش سرشماری، در عمل بروز چنین خطاهایی اجتناب‌ناپذیر نشان می‌دهد. از اینرو سنجش میزان این خطا و به تبع آن تعدیل نتایج سرشماریهای عمومی نفوس از اهمیت بالایی برخوردار می‌باشد.

تحقیقی که گزارش آن را پیش رو دارید به معرفی و بررسی برخی از روشهای برآورد خطای پوشش پرداخته است. گرچه مطالعات انجام گرفته در عرصه عمل امکان بروز نیافته‌اند اما به تبعیت از فازهای سرشماری سایر کشورها امید می‌رود پیش‌آزمون سرشماری عمومی سال ۱۳۸۵ تحت نمونه‌گیری بازشماری (یکی از روشهای معرفی شده در این گزارش) و با رعایت اصول مذکور در این پژوهش مورد آزمون قرار گیرد.

روش نمونه‌گیری بازشماری در برخی سرشماریهای قبلی ایران به منظور برآورد خطای کم‌شماری سرشماری اعمال شده است همچنانکه در روش معرفی شده این گزارش ملاحظه خواهید نمود علاوه بر اختلاف نوع طرح نمونه‌گیری، عوامل بیش‌شماری و در نتیجه برآورد میزان بیش‌شماری در سرشماریهای ایران مورد توجه قرار نمی‌گرفته‌اند. در فصل سوم این گزارش به این عوامل و چگونگی برآورد آنها نیز پرداخته شده است.

به خواننده محترمی که بدون هیچ پیشینه‌ای در زمینه خطاهای پوشش سرشماری به مطالعه این گزارش می‌پردازد توصیه می‌شود حتماً فصل اول گزارش را به عنوان مدخلی به این مبحث پیش از سایر فصول مطالعه نمایند. در این فصل به بررسی علل مطالعه برآورد خطای پوشش سرشماری پرداخته شده است و بر این اساس ۴ روش کلی معرفی شده‌اند که با توجه به الزامات هر یک از این روش‌ها، دو روشی که تا حدودی قابل اجرا در ایران هستند به تفصیل در فصلهای دوم و سوم مورد بررسی قرار گرفته‌اند. در فصل چهارم به برآوردهای زیر ناحیه‌ای خطای پوشش و معرفی چند روش از آن پرداخته‌ایم. از آنجا که در شرایط موجود تنها با استفاده از روش نمونه‌گیری بازشماری

قادر به برآورد قابل اعتمادی از خطای پوشش سرشماری هستیم در فصل خاتمه فازهای مختلف عملیات بازشماری معرفی شده‌اند.

این پژوهش با مدیریت جناب آقای محمدرضا فرید روحانی و با همکاری آقایان یدا... واقعی و حسن رنجی و خانم زهره فلاح محسن‌خانی انجام شده است. همچنین جناب آقای حقیقی نیز به عنوان مشاور این طرح گروه را یاری نمودند. سرکار خانم یزدان‌پناه نیز زحمت تایپ این گزارش را تقبل نمودند. بدین وسیله از زحمات کلیه این عزیزان تشکر می‌شود.

در پایان از خوانندگان محترم تقاضا می‌شود، نظرات اصلاحی خود را در ارتباط با محتوای گزارش، به گروه پژوهشی طرحهای فنی و روشهای آماری پژوهشکده آمار منعکس نمایند.

گروه پژوهشی طرحهای فنی و روشهای آماری

فهرست

فصل اول: تعدیل سرشماری، چرا و چگونه؟

- ۱-۱ مقدمه ۱
- ۲-۱ منابع بروز خطای سرشماری ۲
- ۳-۱ مدلهای سرشماری ۶
- ۴-۱ روشهای تعدیل سرشماری ۸
- ۵-۱ بررسی سرشماریها ۱۴

فصل دوم: برآوردهای جمعیت شناسی خطای پوشش

- ۱-۲ مقدمه ۴۲
- ۲-۲ انواع برآوردهای جمعیت شناختی جمعیت ۴۳
- ۳-۲ برخی روشهای برآورد جمعیت کل کشور به روش تحلیل جمعیت شناختی ۴۴
- ۴-۲ برآورد مؤلفهها ۵۰
- ۵-۲ محدودیتهای برآوردهای جمعیت شناختی ۵۱
- ۶-۲ ارزیابی برآوردهای تحلیل جمعیت شناختی ۵۲

فصل سوم: تعدیل سرشماری به کمک نمونه‌گیری بازشماری

- ۱-۳ مقدمه ۵۷
- ۲-۳ نمونه‌بازشماری (PES) ۵۹
- ۳-۳ مدل سازی ۶۲
- ۴-۳ برآورد DSE تجربی ۷۸

۸۱ ۵-۳ هموارسازی

۸۲ ۶-۳ تجزیه خطای کل

۹۱ ۷-۳ برآورد خطای کل

فصل چهارم: برآوردهای جمعیت نواحی کوچک

۹۳ ۱-۴ مقدمه

۹۴ ۲-۴ روش رگرسیونی

۹۶ ۳-۴ روش تصنعی

۹۸ ۴-۴ برآوردهای تصنعی تحلیل جمعیت شناختی برای نواحی کوچک

۱۰۳ ۵-۴ روش بیز تجربی

۱۰۷ فصل پنجم: خاتمه

۱۱۳ کتابنامه

۱۱۵ ضمیمه

فصل ۱

تعدیل سرشماری، چرا و چگونه؟

۱-۱ مقدمه

سرشماری های نفوس و مسکن هنوز در اکثر کشورهای جهان منبع عمده آمارهای جمعیتی و اقتصادی - اجتماعی است. به علاوه سرشماری ها تنها منبع اطلاعات جزئی جغرافیایی می باشند که با توجه به توسعه کاربردهای سیستم های اطلاعات جغرافیایی (GIS) قوی در تحلیل های فضایی و همچنین توسعه مدیریت شبکه های حمل و نقل تقاضا برای این گونه اطلاعات به سرعت در حال رشد است. بنابراین حتی با وجود بار مالی بالا برای دولت ها هنوز در بیشتر کشورها سرشماری های عمومی انجام می گیرد. اما علیرغم تلاشهای بسیار دست اندرکاران امور سرشماری در کشورهای مختلف برای حصول به استانداردهای بالا در تمامی مراحل عملیات تدارکاتی و میدانی خطاهایی بوجود می پیوندند که نتایج سرشماری را تحت تأثیر قرار می دهند. عمده این خطاها سیستماتیک هستند که بایستی شناسایی و میزان تأثیر آنها در نتایج سرشماری مورد ارزیابی قرار گیرد.

۱-۲ منابع بروز خطای سرشماری

در این بخش به معرفی برخی از منابع خطا که در فرآیند سرشماری رخ می‌دهند اشاره خواهیم داشت.

۱-۲-۱ خطای پوشش

افراد یا واحدهای مسکونی گم شده باعث کم پوششی می‌شوند. واحد مسکونی به علل مختلفی از قلم می‌افتند:

الف) نقص نقشه و بهنگام نبودن آن (هم در موارد مراجعه حضوری و هم ارسال پستی پرسشنامه سرشماری)

ب) عدم تشخیص حدود و ثغور ناحیه سرشماری توسط مأمور

ج) مشخص نبودن واحد مسکونی (مثلاً پلاک نداشتن واحد مسکونی و یا به چشم نیامدن آن واحد)

د) غیرمسکونی قلم داد نمودن واحد مسکونی توسط مأمور سرشماری

ه) عدم مراجعه مأمور به اماکن صعب‌العبور

و) در مواردی که پرسشنامه پست می‌شود شاید در یک آدرس پستی چند خانوار زندگی کنند ولی تنها یک پرسشنامه ارسال شود

ز) خانواده‌هایی که اقامتگاه ثابت ندارند (بیشتر برای خوش نشین‌های روستایی)

افراد نیز از خانوار به علل گوناگونی حذف می‌شوند که غیر از موارد از قلم افتادگی کل خانوار که منجر به حذف افراد آن خانوار از سرشماری می‌شود می‌توان عوامل دیگری را برای حذف افراد از خانوار برشمرد:

الف) حذف فرد توسط خانوار یا انگیزه‌های مختلف (مسئله دار بودن فرد و یا حتی به علت عدم آگاهی پاسخگو، مثلاً نوزادان و پیران کم اهمیت تشخیص داده شده حذف می‌گردند)

ب) حذف افرادی که در مکانهای غیرمتعارف نظیر چادر، کپر و کاروان زندگی می‌کنند

ج) عدم تشخیص فرد توسط مأمور به عنوان عضو خانوار

د) افراد بی‌خانمان در حال حرکت

در مقابل امکان دارد واحد مسکونی و افراد به اشتباه شمرده شوند و یا چند بار به شمار آیند که این امر باعث پوشش اضافی سرشماری می‌گردد. پوشش اضافی در واحدهای مسکونی نیز به علل مختلفی بروز می‌کند.

الف) ابهام در حدود ناحیه سرشماری (که منجر به دو یا چند شماری واحد می‌گردد)

ب) اشتباه مأمور در شمارش واحدهایی که طبق تعریف واحد مسکونی نبوده‌اند

افراد نیز می‌توانند در جریان سرشماری چند بار شمرده شوند و با اصولاً اشتباهی به شمار آیند

الف) افراد با بیش از یک اقامتگاه

ب) شمارش افرادی خارج از جامعه سرشماری (مثلاً خارجی‌ان مقیم و افراد جعلی (شاید به اشتباه))

ج) مردان با بیش از یک همسر

د) شمارش افراد فوت شده قبل از روز سرشماری

ه) شمارش افرادی که پس از روز سرشماری متولد شده‌اند

ب‌طور کلی پوشش اضافی احتمالاً کمتر از کم پوششی رخ می‌دهد و از اینرو محتملاً افراد و اماکن مسکونی کمتر از مقدار واقعی برآورد می‌شوند.

۱-۲-۲ خطاهای بدون پاسخ

در مواردی امکان دارد پاسخهای کاملی در تکمیل پرسشنامه بدست نیاید مثلاً در موارد بسیار نادری فرد پاسخگو (مثلاً رئیس خانوار) از پاسخ دادن یا تکمیل پرسشنامه امتناع می‌کند. در چنین مواردی در ابتدا سعی می‌شود با پیگیری مأمور سرشماری این موارد تکمیل گردد اما در نهایت هنوز اقلام بدون پاسخ وجود خواهند داشت که با جایگذاری آن با موارد مشابه (انتساب) تکمیل می‌گردد ولی در هر صورت خطای بدون پاسخ خواهیم داشت.

۱-۲-۳ خطای پاسخ

حتی اگر پرسشنامه تکمیل شده باشد دلیلی بر صحت تمامی پاسخها نیست. شاید پاسخگو مفهوم سؤال مورد نظر را درک نکرده باشد و یا حتی امکان دارد مأمور سرشماری نیز فاقد چنین درکی بوده باشد. در برخی موارد تعاریف سرشماری دارای ابهام هستند ا به گونه‌ای که عامه مردم آن را به روشنی درک نمی‌کنند مثلاً در سرشماری ایران تعریف خانوار مسأله برانگیز است و در بسیاری موارد برداشت پاسخگویان از خانوار با تعریف سرشماری آن متفاوت است (به بخش (۱-۵-۴) رجوع کنید).

مواردی نیز وجود دارد که یک نفر از اعضای خانوار پاسخگوی تمامی اطلاعات درخواستی مأمور سرشماری می‌شود و شاید برخی از اطلاعات سایر اعضای خانوار را صرفاً بر اساس حدس پاسخ دهد (نقش مأمور در این میان مهم است تا به پاسخهای دقیق برسد و یا لاقلاً آن گزینه بدون پاسخ بماند). به تمام این موارد می‌توان اشتباه مأمور در تکمیل پرسشنامه را نیز اضافه نمود.

۱-۲-۴ خطاهای عملیاتی

برخی از سؤالات در فرم سرشماری تشریحی هستند که طی فرآیندی این پاسخها کدبندی می‌شوند. در این فرآیند نیز امکان بروز خطا وجود دارد مثلاً هنگامیکه پاسخ نامفهوم، ناقص و یا ناخوانا است و یا حتی در مواردی تطبیق جواب تشریحی با کدهای معرفی شده ایجاد اشکال می‌کند.

درباره نقش طراحی پرسشنامه در کاهش خطای پوشش سرشماری به (E.A. Martin, 1995) نگاه کنید.

از دیگر خطاهای عملیاتی بروز خطا به هنگام انتقال اطلاعات پرسشنامه به کامپیوتر است که معمولاً برای کنترل این خطا از دوروش استفاده می‌شود. روش اول ویرایش دقیق (مثلاً کنترل ردیفی اطلاعات) به هنگام ورود اطلاعات است و روش دوم ورود مجدد با اطلاعات نمونه‌ای از فرم‌های سرشماری که قبلاً به کامپیوتر وارد شده‌اند و مقابله آنها با یکدیگر است که در صورت مشاهده ناهماهنگی با فایل اصلی ناهمگنی‌ها بایستی کنترل و اصلاح گردند.

پس از ورود اطلاعات و ویرایش آنها پاسخهای از قلم افتاده و یا ناسازگار شناسایی شده و طی فرآیند انتساب پاسخهای سازگار با سایر اطلاعات پاسخگو و یا پاسخی نظیر موارد مشابه دیگر جایگزین می‌گردد. فراموش نکنیم در فرآیند انتساب پاسخهای «معتبر» و نه لزوماً «صحیح» جایگزین داده‌های گم شده و یا «غیرمعتبر» می‌گردد و لذا احتمالاً دچار خطا خواهیم شد که بایستی مورد توجه قرار گیرد.

۱-۲-۵ خطاهای نمونه‌گیری

در برخی از کشورها در سرشماری نفوس اطلاعات کلی برای تمامی افراد جامعه (تحت عنوان فرم کوتاه سرشماری) پرسیده می‌شود و در کنار آن براساس انتخاب نمونه‌ای تصادفی از خانوارها سوالات تکمیلی در قالب فرم بزرگ سرشماری نیز جمع‌آوری می‌گردد که نتایج این نمونه‌گیری به کل جامعه بسط داده می‌شود و لذا در چنین مواردی با خطاهای نمونه‌گیری نیز مواجه هستیم (برای آشنایی با برخی از سرشماریهای کشورها به بخش ۱-۵-۶ رجوع کنید).

پس از آشنایی با منابع بروز خطا در سرشماریها ارزیابی کیفیت پاسخهای حاصله از سرشماری اهمیت پیدا خواهد نمود. برای ارزیابی کیفیت پاسخهای بدست آمده مطالعات گوناگونی صورت گرفته است مثلاً در سرشماری ۱۹۹۶ کانادا برای هر سؤال، نرخهای پاسخ و نرخ نقص ویرایش محاسبه شده است که این کار در تعیین خطاهای بدون پاسخ و سایر خطاهای بالقوه مفید می‌باشد. همچنین جداول حاصل از اطلاعات سرشماری با برآوردهای پس از سرشماری سرشماری قبل (برای آشنایی با برآوردهای پس از سرشماری به فصل دوم رجوع کنید)، بررسی‌های نمونه‌ای (مثل طرح جاری جمعیت

CPS یا طرح نمونه‌ای نیروی کار) و ثبت‌های اداری مختلف مقایسه شده‌اند که در نتیجه آن امکان اشتباهات سرشماری و یا لااقل تفاوت بین این منابع مختلف مشخص می‌گردد. از دیگر مطالعات انجام گرفته در کانادا مقایسه پاسخهای سرشماری افراد منتخب با اطلاعات سایر منابع در مورد آن فرد می‌باشد. مثلاً برای مشخصات «ثابت» معینی نظیر سن، جنسیت، زبان مادری و محل تولد پاسخهای نمونه‌ای افراد در سرشماری سال ۱۹۹۶ را با پاسخهای نظیر در سرشماری ۱۹۹۱ مقایسه کرده‌اند و از این مقایسه اشتباهات احتمالی شناسایی گردیده است.

۳-۱ مدل‌های سرشماری

با توجه به خطاهای معرفی شده در بخش قبل که عملاً در فرآیند سرشماری امکان بروز می‌یابند مبحث اندازه‌گیری خطای پوشش از اهمیت ویژه‌ای برخوردار می‌شود. همچنانکه در انتهای بخش قبل متذکر شدیم مطالعات گوناگونی برای ارزیابی کیفیت پاسخهای بدست آمده انجام گرفته است. اینک در مورد کنترل خطای پوشش که یکی از عمده‌ترین خطاهای سرشماری است و نتیجه آن بروز کم شماری (خالص) جمعیت می‌باشد بحث خواهیم نمود. در مورد کنترل خطای پوشش دو مدل سرشماری بنام مدل «پوشش کامل» و مدل «پوشش برابری» ارائه شده‌اند.

۱-۳-۱ مدل پوشش کامل

در این مدل هدف این است که با بکارگیری تمامی امکانات و صرف هزینه‌های بسیار زیاد و انجام برنامه‌های منظم و گسترده تقریباً تمام افراد جامعه تحت پوشش سرشماری قرار گیرند و در نتیجه خطای کم شماری و بیش شماری به صفر برسد. طبیعی است که در صورت برقراری پوشش کامل برآورد خطاهای پوشش موردی نخواهد داشت.

۱-۳-۲ مدل برابری یا مدل آماری

طرفداران این مدل مدعی هستند که پوشش کامل امری ابده‌آل است و در عمل وجود خطاهای پوشش سرشماری را اجتناب ناپذیر می‌دانند. البته در مدل برابری ضمن تلاش برای حذف سیستماتیک منابع خطای پیش‌گفته سعی می‌شود با استفاده از منابع اطلاعاتی دیگر برآوردهایی مؤثر و دقیق از میزان کم شماری خالص جمعیت (یعنی تفاضل کم شماری و بیش شماری) بدست آید (توجه کنید که برآورد کم شماری خالص همان برآورد خطای پوشش می‌تواند باشد).

لذا طرفداران مدل برابری معتقدند که نتایج سرشماریهای عمومی نفوس را بایستی با استفاده از منابع اطلاعاتی دیگر تعدیل و یا اصلاح نمود. با توجه به اینکه رابطه متقابل بین تعدیل جمعیت سرشماری و برآوردهای کم شماری خالص وجود دارد یعنی با مشخص شدن هر یک دیگری بدست می‌آید لذا می‌توان برای تعدیل سرشماری به دنبال برآورد کم شماری خالص بود و همچنین با برآورد اصلاحی سرشماری، کم شماری خالص سرشماری را برآورد نمود یعنی با داشتن N_c ، رابطه هم‌ارزی زیر برقرار است.

$$\hat{N} \Leftrightarrow \text{برآوردی خالص کم شماری}$$

تذکر: در برخی از کشورهای جهان (بطور عمده اروپایی) افزایش بار مالی سرشماری و همچنین فشارهایی که باعث شده است تا دست‌اندرکاران امور سرشماری از قبول مسئولیت اشتباهات سرشماری شانه خالی کنند، تلاشی را جهت ارائه راه‌حل‌های بهتر در زمینه جمع‌آوری اطلاعات موردنظر در سرشماری به همراه داشته است. این تلاشها عمدتاً منجر به افزایش استفاده از ثبت‌های اداری (آمارهای ثبتی) و توسعه پردازش خودکار اطلاعات در مراحل بازمینی، ویرایش، کدگذاری و ... در اتحادیه اروپایی شده است.

۴-۱ روشهای تعدیل سرشماری

چهار فن اصلی برای برآورد خطای پوشش سرشماری و در نتیجه تعدیل نتایج سرشماری مورد استفاده قرار می‌گیرند که عبارتند از:

- ۱- مقابله ثبت اداری (ARM)^۱
- ۲- کنترل ثبت معکوس (RRC)^۲
- ۳- تحلیل جمعیت شناختی (DA)^۳
- ۴- نمونه‌گیری بازشماری (PES)^۴

۱-۴-۱ مقابله ثبت اداری (ARM)

ثبت های اداری را می‌توان در هریک از روشهای تعدیل بکار گرفت حال یا به تنهایی و یا در ترکیب با سایر برآوردها و مقابله با نتایج سرشماری. یقیناً برای نیل به اصلاح نتایج سرشماری براساس مقابله ثبت‌های اداری بایستی میزان دقت، فراگیری و بهنگام بودن این آمارها در حد بالایی باشد.

۲-۴-۱ کنترل ثبت معکوس (RRC):

در این روش افراد نمونه‌ای که انتظار داریم در زمان سرشماری در کشور بوده باشند را ردگیری می‌کنیم برای انتخاب این افراد نمونه از چندین فهرست مختلف بطور همزمان سود می‌جوئیم که عبارتند از:

الف) افرادی که در سرشماری قبل شمارش شده‌اند

ب) افرادی که در حد فاصل بین دو سرشماری وارد کشور شده‌اند (براساس آمارهای موجود در اداره

مهاجرت)

Adinistrative Record Matching^۱

Reverse Record Check^۲

Demographic Analysis^۳

Post Enumeration Survey^۴