دانشگاه علامه طباطبایی
دانشکده‌ی اقتصاد
گروه آمار، ریاضی و کامپیوتر

پایان‌نامه برای دریافت درجه‌ی کارشناسی‌ارشد **آمار اقتصادی و اجتماعی**

عنوان

براورد کم‌شماری در سرشماری‌های عمومی نفوس و مسکن

پژوهش‌گر

سپیده مسافری

استاد راهنما

دکتر حمیدرضا نواب‌پور

استاد مشاور

دکتر روشنک علی‌اکبری صبا

دی ۱۳۹۰

بسم الله الرحمن الرحيم

کلیه‌ی حقوق مادی و معنوی اعم از چاپ و تکثیر، نسخه‌برداری، ترجمه، اقتباس و ... از این پایان‌نامه

برای دانشگاه علامه طباطبایی محفوظ است. نقل مطالب با ذکر منبع مانعی ندارد.

تأیید پایان‌نامه‌ی کارشناسی‌ارشدتوسط دانشجو

عنوان پایان‌نامه: براورد کم‌شماری در سرشماری‌های عمومی نفوس و مسکن

نام دانشجو: سپیده مسافری

شماره‌ی دانشجویی: ۸۸۱۲۵۱۱۶۱۱۵

استاد راهنما: دکتر حمیدرضا نواب‌پور

این‌جانب **سپیده مسافری** دانشجوی کارشناسی‌ارشد رشته‌ی **آمار اقتصادی و اجتماعی** دانشکده‌ی اقتصاد دانشگاه علامه طباطبایی گواهی می‌نمایم پژوهش‌های ارایه شده در پایان‌نامه با عنوان مذکور توسط شخص این‌جانب انجام شده است و درستی مطالب نگارش یافته مورد تأیید می‌باشد. همچنین گواهی می‌نمایم مطالب مندرج در پایان‌نامه تاکنون برای دریافت هیچ نوع مدرک یا امتیازی توسط این‌جانب یا فرد دیگری در هیچ کجا ارایه نشده است و در نگارش متن پایان‌نامه شیوه‌ی نگارش مصوب دانشکده‌ی اقتصاد را به‌طور کامل رعایت نموده‌ام. چنان‌چه در هر زمان خلاف آن‌چه گواهی نموده‌ام مشاهده گردد خود را از آثار حقیقی و حقوقی ناشی از دریافت مدرک کارشناسی‌ارشد محروم می‌دانم و هیچ‌گونه ادعایی نخواهم داشت.

تاریخ: **امضای دانشجو:**

اهدا یه

تقدیم به پدر و مادر عزیزم،

به پاس حمایت ها و محبت های بی دریغشان

سپاس‌گزاری...

سپاس خداوندی را که در احدیتش یگانه و در واحدیتش یکتاست و ثنای او را که به فضل و رحمت خویش به راه هدایتش راهبرمان گشته و به عنایت، نجات و رسیدن به درجات برای ما قرار داده است. شایسته است از زحمت‌های بی‌دریغ استاد راهنمای بزرگوارم، جناب آقای دکتر حمیدرضا نواب‌پور، تشکر و قدردانی کنم که در تمام مراحل پشتیبان و مشوق بنده بودند و با صبر و بردباری مرا یاری دادند.

از استاد مشاورم سرکار خانم دکتر روشنک علی‌اکبری صبا به پاس راهنمایی‌هایشان، تشکر می‌کنم. همچنین از جناب آقای دکتر نادر نعمت‌الهی و جناب آقای دکتر فرزاد اسکندری که زحمت داوری این اثر را به عهده داشتند سپاس‌گزارم. از پژوهشکده‌ی آمار ایران نیز در خصوص حمایت‌های مادی و معنوی از این پایان‌نامه کمال تشکر را دارم.

در پایان، از خانواده‌ی مهربانم به‌خصوص پدر و مادر عزیزتر از جانم که همواره یار و یاورم بوده و هستند، قدردانی می‌کنم. بر دستانشان بوسه می‌زنم و امید دارم که دعای خیرشان همواره بدرقه‌ی راهم باشد.

امیدوارم بتوانم از عهده‌ی ادای حق این عزیزان برآیم.

دی ۹۰

فهرست مطالب

فهرست مطالب ب

فهرست جدول‌ها ج

نمادها و علایم اختصاری چ

۱ آشنایی ۱

۱.۱ مقدمه ۱

۲.۱ تعریف مفاهیم و واژه‌های اساسی ۴

۱.۲.۱ آلگوریتم دمینگ-استفان ۴

۲.۲.۱ آلگوریتم نیوتون-رافسون ۶

۳.۲.۱ آمارگیری پساسرشماری ۷

۴.۲.۱ اثر خوشه‌بندی ۷

۵.۲.۱ اقامتگاه گروهی (مؤسسه‌ای) ۸

۶.۲.۱ براورد به روش گشتاوری ۸

۷.۲.۱ براورد ماکسیمم درست‌نمایی ۸

۸.۲.۱ براورد ماکسیمم درست‌نمایی مقید ۹

۹.۲.۱ پساطبقه‌بندی ۹

۱۰.۲.۱ جدول ناکامل ۹

۱۱.۲.۱ خانوار معمولی ساکن ۱۰

۱۲.۲.۱ روش امتیازدهی ۱۰

۱۳.۲.۱ روش دلتا ۱۱

۱۴.۲.۱ روش‌های تکراری برای براورد واریانس ۱۱

۱۵.۲.۱ فضای ستونی ماتریس ۱۲
۱۶.۲.۱ مدل راش ۱۲
۱۷.۲.۱ مدل شبه‌تقارنی ۱۲
۱۸.۲.۱ مدل شبه‌لگ‌خطی ۱۵
۱۹.۲.۱ مقایسه‌ی مقید خطایی ۱۵
۲۰.۲.۱ نسبت جنسیتی ۱۵
۳.۱ مرور نوشتگان ۱۵
۴.۱ هدف پژوهش ۱۸
۵.۱ چشم‌انداز فصل‌های آینده ۱۸

۲ روش‌های اجرای سرشماری جمعیتی ۱۹
۱.۲ مقدمه ۱۹
۲.۲ سرشماری سنتی ۲۱
۳.۲ سرشماری ثبتی مبنا ۲۲
۴.۲ ترکیب سرشماری سنتی و سرشماری ثبتی مبنا ۲۳
۵.۲ سرشماری ثبتی مبنا با آمارگیری‌های نمونه‌ای ۲۴
۶.۲ سرشماری سنتی با روزامدسازی سالانه‌ی ویژگی‌ها ۲۴
۷.۲ سرشماری غلتان ۲۵
۱.۷.۲ تجربه‌ی فرانسه در اجرای سرشماری غلتان ۲۶
۸.۲ چکیده‌ی فصل ۲۹

۳ روش‌های براورد کم‌شماری خالص ۳۰
۱.۳ مقدمه ۳۰
۲.۳ تحلیل جمعیت‌شناختی ۳۱
۱.۲.۳ اریبی جمعیت براوردشده توسط تحلیل جمعیت‌شناختی ۳۳
۲.۲.۳ نسبت جنسیتی در تحلیل جمعیت‌شناختی ۳۴
۳.۲.۳ براورد کم‌شماری خالص با استفاده از تحلیل جمعیت‌شناختی ۳۵
۳.۳ نظام دوگان ۳۷
۱.۳.۳ رویکردهای آمارگیری پساسرشماری ۴۳

۲.۳.۳ اریبی همبستگی در نظام دوگان ۴۷
۱.۲.۳.۳ تعدیل اریبی همبستگی در سطح ملی ۴۹
۲.۲.۳.۳ تعدیل اریبی همبستگی در سطح پساطبقه‌ها ۴۹
۳.۳.۳ براورد کم‌شماری خالص با استفاده از نظام دوگان ۵۳
۴.۳ نظام سه‌گان ۵۴
۱.۴.۳ براورد جمعیت واقعی در نظام سه‌گان بر اساس وابستگی میان نمونه‌ها . . . ۵۶
۲.۴.۳ براورد جمعیت واقعی در نظام سه‌گان بر اساس گیرپذیری ناهمگن میان افراد . ۶۱
۳.۴.۳ براورد کم‌شماری خالص با استفاده از نظام سه‌گان ۶۹
۵.۳ رویکرد بیز تجربی ۷۰
۱.۵.۳ انواع روش‌های براورد پارامترهای ماتریس واریانس کوواریانس ۷۳
۱.۱.۵.۳ براورد ماکسیمم درست‌نمایی ۷۴
۲.۱.۵.۳ براورد به روش گشتاوری ۷۶
۳.۱.۵.۳ براورد ماکسیمم درست‌نمایی مقید ۷۷
۲.۵.۳ پیشگویی کم‌شماری خالص با استفاده از بیز تجربی ۷۹
۶.۳ چکیده‌ی فصل ۸۰

۴ مطالعه‌ی کاربردی **۸۱**

۱.۴ مقدمه ۸۱
۲.۴ نحوه‌ی براورد کم‌شماری در ایران ۸۱
۱.۲.۴ براورد کم‌شماری در سرشماری ۱۳۷۵ . . . . . ۸۲
۲.۲.۴ براورد کم‌شماری در سرشماری ۱۳۸۵ . . . . . ۸۴
۳.۲.۴ مقایسه‌ی نحوه‌ی براورد کم‌شماری در سرشماری‌های ۱۳۷۵ و ۱۳۸۵ . . . . ۸۹
۳.۴ نحوه‌ی براورد کم‌شماری در ایالت‌های متحده ۹۰
۱.۳.۴ براورد کم‌شماری در سرشماری ۲۰۰۰ . . . . . ۹۰
۲.۳.۴ براورد کم‌شماری در سرشماری ۲۰۱۰ . . . . . ۹۸
۳.۳.۴ مقایسه‌ی نحوه‌ی براورد کم‌شماری در سرشماری‌های ۲۰۰۰ و ۲۰۱۰ . . . . ۱۰۲
۴.۴ مقایسه‌ی عمل‌کرد ایران و ایالت‌های متحده ۱۰۴
۵.۴ ارایه‌ی پیش‌نهادی به منظور براورد کم‌شماری در ایران ۱۰۴

مرجع‌ها **۱۰۸**

واژه‌نامه‌ی فارسی به انگلیسی **۱۱۴**

پیوست الف برهان برخی رابطه‌ها **۱۲۵**

فهرست جدول‌ها

۱.۳ جدول پیش‌ایندی $۲^۲$ با توجه به دو منبع اطلاعاتی ۴۰

۲.۳ جدول پیش‌ایندی $۲^۳$ با توجه به سه منبع اطلاعاتی ۵۵

۱.۴ حالت‌های مختلف در عملیات تطبیق خانوارها ۸۳

۲.۴ حالت‌های مختلف در عملیات تطبیق افراد ۸۴

۳.۴ وضعیت تطبیق خانوارها در مرحله‌ی اول تطبیق ۸۶

۴.۴ وضعیت تطبیق افراد در مرحله‌ی اول تطبیق ۸۷

۵.۴ وضعیت تطبیق افراد و خانوارها در مرحله‌ی دوم تطبیق ۸۷

۶.۴ وضعیت افراد P نمونه و E نمونه بعد از عملیات تطبیق رایانه‌ای و دفتری ۹۳

۷.۴ وضعیت نهایی افراد P نمونه و E نمونه بعد از عملیات پیگیری ۹۳

۸.۴ اطلاعات مدل رگرسیون لوژستیکی ۱۰۲

نمادها و علایم اختصاری

تعریف	نماد
اثر	tr
دترمینان	det
درجه آزادی	df
متناسب است با	$\propto$
مشتق	∂
آمارگیری پساسرشماری	PES
ارزیابی پوشش و درستی	ACE
اندازه‌گیری پوشش سرشماری	CCM
اردوگاه سرخ‌پوستان امریکایی	AIR
تحلیل جمعیت‌شناختی	DA
فایل آدرس پایه‌ای	$MAF/TIGER$
فایل آدرس پایه‌ای روزامدشده	UCM
فایل ادیت‌نشده‌ی سرشماری	CUF
فهرست مستقل ارزیابی پوشش و درستی	$ACEIL$
فهرست مستقل اندازه‌گیری پوشش سرشماری	$CCMIL$
نظام دوگان	DS

چکیده

سرشماری‌های نفوس و مسکن جامع‌ترین منبع اطلاعاتی در ارتباط با ویژگی‌های جمعیتی و واحدهای مسکونی هستند. از هدف‌های اصلی سرشماری‌های نفوس، فراهم کردن اطلاعات لازم برای سیاست‌گذاری‌ها، برنامه‌ریزی‌ها، تهیه‌ی چارچوب‌های آماری و ... هستند. بنابر این کشورها با توجه به امکانات دردسترس، هزینه و ویژگی‌های جمعیتی، روشی مناسب را برای اجرای سرشماری در نظر می‌گیرند.

خطاهای غیرنمونه‌گیری بسیاری در سرشماری‌ها رخ می‌دهند. یکی از این خطاها، خطای پوشش است. خطای پوشش بر دو نوع کم‌شماری و بیش‌شماری است، که معمولاً خطای کم‌شماری بیش‌تر رخ می‌دهد. بنابر این براورد کم‌شماری خالص از اهمیت خاصی برخوردار است. لذا بسیاری از کشورها پس از اتمام سرشماری‌ها برای بررسی میزان درستی شمارش‌ها، با به کارگیری روشی خاص کم‌شماری خالص را براورد می‌کنند.

در این پایان‌نامه، ابتدا ضرورت اجرای سرشماری بیان می‌شود. انواع روش‌های معمول اجرای سرشماری در کشورها توضیح داده می‌شوند. سپس خطای پوشش و علل رخ‌داد آن معرفی می‌شوند. ضرورت بررسی خطای پوشش مطرح شده و نحوه‌ی براورد آن توضیح داده می‌شود. از آن‌جا که براورد خطای پوشش خالص از اهمیت بیش‌تری برخوردار است، لذا با توجه به روش‌هایی که مطرح خواهند شد، تلاش می‌شود ابتدا جمعیت واقعی براورد شود. سپس براورد خطای پوشش خالص از اختلاف شمارش سرشماری از جمعیت واقعی براوردشده، به‌دست می‌آید. به‌طور معمول این اختلاف مقدار مثبتی دارد که بیانگر کم‌شماری خالص است.

سرانجام به‌عنوان یک کاربرد، نظام دوگان به‌عنوان کاربردی‌ترین روش در بسیاری از کشورها، در ایران و ایالت‌های متحده مطرح خواهد شد. برای این منظور، نحوه‌ی اجرای آمارگیری پساسرشماری و استفاده از نظام دوگان برای سال‌های ۱۳۷۵ و ۱۳۸۵ در ایران و برای سال‌های ۲۰۰۰ و ۲۰۱۰ در ایالت‌های متحده در نظر گرفته شده است. بعد از بیان نحوه‌ی اجرای فعالیت‌ها در این دو کشور و مقایسه‌ها، پیش‌نهادی به منظور براورد مناسب‌تر کم‌شماری خالص در ایران ارایه خواهد شد.

واژگان کلیدی. *آمارگیری پساسرشماری؛ جمعیت واقعی؛ خطای پوشش؛ خطای کم‌شماری خالص؛ سرشماری؛ نظام دوگان.*

فصل ۱
آشنایی

۱.۱ مقدمه

برای برنامه‌ریزی در سطح‌های مختلف یک کشور دانستن جمعیت مربوط به آن الزامی است. در بسیاری از موقعیت‌ها با توجه به کمبود اطلاعات آماری، واحدهای تولیدکننده‌ی آمار برای تولید داده‌های مورد نیاز، از آمارگیری‌هایی همچون نمونه‌گیری و سرشماری استفاده می‌کنند. در سرشماری از تمام افراد یا واحدهای جمعیت مورد نظر اطلاع‌گیری می‌شود. در صورتی‌که سرشماری در سراسر کشور به‌عمل آید، به آن سرشماری عمومی می‌گویند. سرشماری، جامع‌ترین منبع اطلاعاتی در ارتباط با جمعیت و خانوار است که اطلاعاتی در زمینه‌ی اندازه، ساختار و ویژگی‌های جمعیت ارایه می‌دهد. با استفاده از سرشماری می‌توان امید زندگی (متوسط طول عمر) مردم یک کشور را محاسبه کرد و فهمید جمعیت کشور جوان است یا پیر؟ آیا تعداد بیمارستان‌های کشور کافی هستند؟ چند درصد از افراد دارای بیماری‌های خاص هستند؟ آیا تعداد مدرسه‌های کشور کافی هستند؟

اطلاعات سرشماری برای تحلیل مسایل جمعیت‌شناختی، اجتماعی، اقتصادی و گروه‌های نادر همچون اقلیت‌های قومی، مذهبی و ... بسیار مؤثر هستند. همچنین دولت‌ها، مرکزهای بهداشتی و تمام سازمان‌هایی که وظیفه‌ی تخصیص منبع‌های کشور، برنامه‌ریزی‌ها و سیاست‌گذاری‌ها را بر عهده دارند از این اطلاعات

استفاده می‌کنند. علاوه بر این، سرشماری چارچوب‌های آماری لازم را برای آمارگیری‌های نمونه‌ای مهیا می‌سازد. همچنین اطلاعات مورد نیاز را در سطح کوچک‌ترین واحد جغرافیایی کشور تهیه می‌کند. در بسیاری از کشورها، سرشماری نفوس با سرشماری مسکن همراه است تا اطلاعاتی درباره‌ی ویژگی‌های محل سکونت افراد به‌دست آید. ایران هم جزء یکی از این کشورهاست.

از آن‌جا که اجرای سرشماری اجتناب‌ناپذیر است، لذا کشورها پیوسته به دنبال روش‌هایی هستند تا با توجه به شرایط کشور پوشش بیش‌تری از جمعیت را ارایه کنند. روش‌های معمول عبارت‌اند از:

- سرشماری سنتی،
- سرشماری ثبتی مبنا،
- ترکیب سرشماری سنتی و سرشماری ثبتی مبنا،
- سرشماری ثبتی مبنا با آمارگیری‌های نمونه‌ای،
- سرشماری سنتی با روزامدسازی سالانه‌ی ویژگی‌ها، و
- سرشماری غلتان.

در سرشماری‌ها، خطاهای غیرنمونه‌گیری بسیاری همچون خطای پوشش، خطای محتوا و خطای عملیاتی رخ می‌دهند. گاهی مشکلات در حین سرشماری باعث ایجاد خطا در شمارش سرشماری می‌شوند که خطای پوشش نامیده شده و در میان دیگر خطاها اهمیت بیش‌تری دارد. زیرا از هدف‌های اصلی سرشماری، شمارش درست و کامل جمعیت است. خطای پوشش بر دو نوع است: کم‌شماری و بیش‌شماری.

در سرشماری ممکن است برخی از واحدهای جامعه‌ای ازقلم‌بیفتند یا برقراری تماس با تمام واحدهای جامعه‌ای ممکن نباشد. حتی در صورت برقراری تماس، به‌دست آوردن اطلاعات از برخی واحدها دشوار است. این عدم‌پوشش منجر به کم‌شماری می‌شود. در حقیقت کم‌شماری به معنای شمارش نشدن برخی از افراد جمعیت است. گاهی در سرشماری‌ها، بیش‌شماری نیز رخ می‌دهد. بیش‌شماری به معنای شمارش برخی از افراد بیش از یک بار یا شمارش افراد خارج از جمعیت است.

بنابر این، مشخص کردن جمعیتی که جزء جامعه‌ی هدف سرشماری هستند، اهمیت زیادی دارد. با توجه به روش اجرای سرشماری و تعریف‌هایی که یک کشور در نظر می‌گیرد، می‌توان شایستگی شمارش شدن فرد را در هنگام سرشماری مورد بررسی قرار داد. به‌عنوان مثال، ممکن است برای یک کشور، مهاجری که به‌طور قانونی و دایمی در کشور دیگری در حال زندگی است، شایستگی شمارش در سرشماری را داشته باشد.

اما نوزادی که روز بعد از سرشماری به دنیا آمده، شایستگی شمارش در سرشماری را نداشته باشد. تعیین شایستگی یا عدم‌شایستگی افراد در هنگام سرشماری، در کاهش خطای پوشش مؤثر است.

بعد از اتمام سرشماری، کشورها با به کارگیری روش‌های مناسب سعی می‌کنند تا جمعیت واقعی را به‌درستی براورد کنند. سپس از اختلاف شمارش سرشماری از جمعیت واقعی براوردشده، خطای پوشش خالص را براورد کنند. در نتیجه خطای پوشش خالص از رابطه‌ی زیر براورد می‌شود:

شمارش سرشماری - براورد جمعیت واقعی = براورد خطای پوشش خالص

به‌طور معمول در بسیاری از سرشماری‌ها، خطای کم‌شماری بیش از خطای بیش‌شماری رخ می‌دهد و این اختلاف مقدار مثبتی دارد که بیانگر کم‌شماری خالص در هنگام سرشماری است. اما گاهی این اختلاف مقدار منفی دارد که بیانگر بیش‌شماری خالص است.

به‌عبارت دیگر، می‌توان خطای پوشش خالص را به‌صورت اختلاف مقدار بیش‌شماری از مقدار کم‌شماری نیز تعریف کرد، که به‌صورت زیر براورد می‌شود:

براورد بیش‌شماری - براورد کم‌شماری = براورد خطای پوشش خالص

بنابر این، خطای پوشش ناخالص به‌صورت مجموع مقدار کم‌شماری و بیش‌شماری است که به‌صورت زیر براورد می‌شود:

براورد بیش‌شماری + براورد کم‌شماری = براورد خطای پوشش ناخالص

اما به‌طور معمول از اختلاف مقدار بیش‌شماری از مقدار کم‌شماری برای براورد خطای پوشش خالص استفاده نمی‌شود، زیرا بسیاری از روش‌ها توانایی ارایه‌ی براوردهایی مجزا را برای این دو خطا ندارند. بنابر این دو رابطه‌ی اخیر تنها به منظور درک بیش‌تر مطلب مطرح شده‌اند. با استفاده از روش‌هایی که در ادامه مطرح خواهند شد، جمعیت واقعی براورد می‌شود. سپس از اختلاف شمارش سرشماری از جمعیت واقعی براوردشده، خطای پوشش خالص براورد می‌شود.

معمولاً به دلیل وقوع بیش‌تر خطای کم‌شماری نسبت به خطای بیش‌شماری، این اختلاف مقدار مثبتی دارد که بیانگر کم‌شماری خالص است. بنابر این در بسیاری از مبحث‌های مربوط به ارزیابی پوشش سرشماری از عنوان کم‌شماری استفاده می‌شود. همچنین در کشورهایی که اطلاعات به‌صورت مصاحبه‌ی حضوری از افراد گرداوری می‌شوند، اغلب خطای کم‌شماری بیش‌تر از خطای بیش‌شماری رخ می‌دهد.

عامل‌های مختلف سیاسی، فرهنگی، قومی و نقص نقشه‌های آماری باعث می‌شوند که برخی افراد در سرشماری اظهار نشوند و جمعیتی کم‌تر از آن‌چه که هست، شمارش شود. کم‌شماری باعث می‌شود برخی از افرادی که در جامعه‌ی هدف وجود دارند در چارچوب‌های نمونه‌گیری قرار نگیرند. از آن‌جا که این چارچوب‌ها در طرح‌های آمارگیری استفاده می‌شوند، بخشی از جامعه‌ی هدف در هر آمارگیری حضور ندارد.

این امر می‌تواند سبب اریبی گزینش در برخی نمونه‌گیری‌ها شود. علاوه بر این، کم‌شماری سیاست‌گذاری‌ها و برنامه‌ریزی‌های یک کشور را تحت‌تأثیر قرار می‌دهد. مثلاً شهری که می‌تواند به شهرستان تبدیل شود همچنان شهر باقی می‌ماند و سهم کم‌تری از منبع‌های مالی کشور و نمایندگان مجلس را به‌خود اختصاص می‌دهد. از این‌رو تلاش بسیاری برای اجرای دقیق سرشماری‌ها و براورد کم‌شماری صورت می‌گیرد.

روش‌های مختلفی با توجه به شرایط کشورها برای براورد کم‌شماری خالص وجود دارند. اما روش‌هایی که به‌طور معمول برای براورد کم‌شماری خالص به کار می‌روند، عبارت‌اند از:

- تحلیل جمعیت‌شناختی،
- نظام دوگان،
- نظام سه‌گان، و
- رویکرد بیز تجربی.

در میان این روش‌ها، تحلیل جمعیت‌شناختی و نظام دوگان (با استفاده از اطلاعات آمارگیری پساسرشماری و سرشماری) از اهمیت زیادی برخوردار هستند و بسیاری از کشورها نیز از این دو روش به‌خصوص نظام دوگان استفاده می‌کنند.

براورد کم‌شماری نیز با خطا همراه است و نتیجه‌های آن معمولاً برای تعدیل شمارش سرشماری استفاده نمی‌شود (فریدمن و واچر، ۲۰۰۱)، بلکه از نتیجه‌های کم‌شماری برای براوردهای جمعیتی پس از سرشماری، بهبود تصمیم‌گیری‌ها، برنامه‌ریزی‌ها و نقشه‌های آماری استفاده می‌شود. کشورهایی که معمولاً تعدیل را روی نتیجه‌های سرشماری اعمال می‌کنند، معتقدند که روش‌های براورد جمعیت واقعی و در نتیجه کم‌شماری دقیق‌تر از سرشماری انجام می‌شوند. با این حال، تعدیل یا عدم‌تعدیل نتیجه‌های سرشماری بیش‌تر یک تصمیم‌گیری سیاسی و مدیریتی است تا یک تصمیم‌گیری آماری.

۲.۱ تعریف مفاهیم و واژه‌های اساسی

در این قسمت برخی مفهوم‌های پایه‌ای به کار رفته در این پایان‌نامه معرفی می‌شوند.

۱.۲.۱ آلگوریتم دمینگ-استفان

گاهی اوقات روش تحلیلی برای یافتن براوردهای ماکسیمم درست‌نمایی مدل‌های لگ‌خطی وجود ندارد. در این موردها می‌توان از روشی که توسط دمینگ و استفان (۱۹۴۰) ارایه شد، استفاده کرد. این آلگوریتم در ارتباط با مجموع‌های حاشیه‌ای یک جدول پیش‌ایندی است و تنها برای به‌دست آوردن براوردهای ماکسیمم

درست‌نمایی در مدل‌های لگ‌خطی و شبه‌لگ‌خطی (در مفهوم ۱۶.۲.۱ تعریف شده است) استفاده می‌شود. اما برای مدل‌های شبه‌لگ‌خطی، براوردهای ماکسیمم درست‌نمایی تنها برای خانه‌های مشاهده‌شده محاسبه می‌شوند.

آلگوریتم دمینگ-استفان برای شروع از کوچک‌ترین شمارشی که یک خانه در جدول می‌تواند به‌خود اختصاص دهد، شروع می‌کند. سپس یک چرخه‌ی تکراری که شامل s گام است، آغاز می‌شود. مقدار s بستگی به تعداد بالاترین اثرها و اثرهای اصلی موجود در مدل دارد. این چرخه تا جایی ادامه می‌یابد که یک دقت کافی حاصل شود، سپس براوردهای ماکسیمم درست‌نمایی به‌دست می‌آیند.

برای یک جدول پیش‌ایندی $I \times J \times \ldots \times K \times L$، $x_{ij\ldots kl}$ شمارش مشاهده‌شده‌ی خانه‌ها، $m_{ij\ldots kl}$ شمارش موردانتظار خانه‌ها و $\hat{m}_{ij\ldots kl}$ براورد ماکسیمم درست‌نمایی شمارش موردانتظار خانه‌هاست. نحوه‌ی عمل‌کرد آلگوریتم به‌صورت زیر است:

گام صفر: برای خانه‌هایی از جدول که مشاهده شده‌اند $\hat{m}^{(\cdot)}_{ij\ldots kl} = ۱$ و برای خانه‌هایی از جدول ناکامل که مشاهده نشده‌اند $\hat{m}^{(\cdot)}_{ij\ldots kl} = ۰$ است. سپس چرخه‌ی تکراری آغاز می‌شود. برای آگاهی بیش‌تر از گام‌های این چرخه فرض کنید، مدل اشباع‌شده‌ی زیر با تمام اثرها وجود داشته باشد:

$$\log m_{ij\ldots kl} = \mu + \alpha_i + \beta_j + \ldots + \gamma_k + \lambda_l + (\alpha\beta)_{ij} + \ldots + (\gamma\lambda)_{kl} + \ldots + (\alpha\beta\ldots\gamma\lambda)_{ij\ldots kl}$$

در این مدل μ اثر میانگین سراسری، α_i، β_j، γ_k، ... و λ_l اثرهای اصلی و $(\alpha\beta)_{ij}$، $(\gamma\lambda)_{kl}$، ... و $(\alpha\beta\ldots\gamma\lambda)_{ij\ldots kl}$ اثرهای متقابل هستند. بالاترین اثر $(\alpha\beta\ldots\gamma\lambda)_{ij\ldots kl}$ است و شامل تمام اثرهای اصلی مدل می‌باشد. بنابر این $s = ۱$ و چرخه تنها یک گام دارد، که به‌صورت زیر است:

$$\hat{m}^{\nu}_{ij\ldots kl} = x_{ij\ldots kl}$$

با توجه به این چرخه، ν مقدارهای طبیعی $(\nu = ۱, ۲, ۳, \ldots)$ را اختیار می‌کند. چون مدل اشباع‌شده است، با تکرار چرخه جواب‌های یکسانی به‌دست می‌آید.

اگر اثر متقابل $(\alpha\beta\ldots\gamma\lambda)_{ij\ldots kl}$ از مدل لگ‌خطی مطرح‌شده در بالا حذف شود، آن‌گاه گام صفر تحت این مدل همانند قبل خواهد بود و چرخه‌ی تکراری به‌صورت زیر است:

گام اول:
$$\hat{m}^{s\nu-(s-۱)}_{ij\ldots kl} = \frac{\hat{m}^{(s\nu-s)}_{ij\ldots kl}\; x_{ij\ldots k+}}{\hat{m}^{(s\nu-s)}_{ij\ldots k+}}$$

گام دوم:
$$\hat{m}^{s\nu-(s-۲)}_{ij\ldots kl} = \frac{\hat{m}^{(s\nu-(s-۱))}_{ij\ldots kl}\; x_{ij\ldots +l}}{\hat{m}^{(s\nu-(s-۱))}_{ij\ldots +l}}$$

و ... گام sام:
$$\hat{m}^{s\nu}_{ij\ldots kl} = \frac{\hat{m}^{(s\nu-۱)}_{ij\ldots kl}\; x_{+j\ldots kl}}{\hat{m}^{(s\nu-۱)}_{+j\ldots kl}}$$

در این چرخه $x_{ij...kl} = \sum_{l} x_{ij...kl}$، $x_{ij...k+} = \sum_{k} x_{ij...kl}$، ... و $x_{+j...kl} = \sum_{i} x_{ij...kl}$ است و چرخه هنگامی متوقف می‌شود که از دقت کافی برخوردار باشد. برای این منظور باید جواب به‌دست آمده از گام sام چرخه‌ی νام به اندازه‌ی ε، $\varepsilon > ۰$، از گام sام چرخه‌ی $(\nu - ۱)$ام اختلاف داشته باشد.

اگر در مدل لگ‌خطی تمام اثرهای متقابل حذف شوند، مدل لگ‌خطی به‌صورت زیر خواهد بود:

$$\log m_{ij...kl} = \mu + \alpha_i + \beta_j + ... + \gamma_k + \lambda_l$$

که با توجه به این مدل، چرخه‌ی تکراری به‌صورت زیر است:

گام اول:
$$\hat{m}_{ij...kl}^{s\nu-(s-۱)} = \frac{\hat{m}_{ij...kl}^{(s\nu-s)} \quad x_{i+...++}}{\hat{m}_{i+...++}^{(s\nu-s)}}$$

گام دوم:
$$\hat{m}_{ij...kl}^{s\nu-(s-۲)} = \frac{\hat{m}_{ij...kl}^{(s\nu-(s-۱))} \quad x_{+j...++}}{\hat{m}_{+j...++}^{(s\nu-(s-۱))}}$$

و ... گام sام:
$$\hat{m}_{ij...kl}^{s\nu} = \frac{\hat{m}_{ij...kl}^{(s\nu-۱)} \quad x_{++...+l}}{\hat{m}_{++...+l}^{(s\nu-۱)}}$$

در این چرخه $x_{i+...++} = \sum_{j...kl} x_{ij...kl}$، $x_{+j...++} = \sum_{i...kl} x_{ij...kl}$، ... و $x_{++...+l} = \sum_{ij...k} x_{ij...kl}$ است و چرخه هنگامی متوقف می‌شود که از دقت کافی برخوردار باشد.

۲.۲.۱ آلگوریتم نیوتون-رافسون

هنگامی‌که روش تحلیلی برای یافتن ریشه‌ی تابع دلخواه $f(x)$ وجود ندارد، می‌توان از روش‌های عددی برای یافتن ریشه استفاده کرد. یکی از روش‌های عددی معمول برای یافتن ریشه‌ی تابع، استفاده از آلگوریتم نیوتون-رافسون بر اساس بسط تیلور است.

فرض کنید تابع $f(x)$ بر بازه‌ی $[a,b]$ دو بار به‌طور پیوسته مشتق‌پذیر باشد. اگر $c \in [a,b]$ یک تقریب به p، ریشه‌ی تابع، باشد به‌طوری‌که $f'(c) \neq ۰$ و $|p-c|$ کوچک باشد، آن‌گاه بسط تیلور مرتبه‌ی اول $f(x)$ حول c به‌صورت زیر است:

$$f(x) = f(c) + (x-c)f'(c) + o(|x-c|) \quad x \to c \qquad (۱.۲.۱)$$

در رابطه‌ی (۱.۲.۱)، $o(.)$ بدین معناست که اگر x_n هر دنباله‌ای باشد که $x_n \to c$ میل کند و دنباله‌های a_n و b_n به‌صورت زیر تعریف شوند:

$$a_n = f(x_n) - f(c) - (x_n - c)f'(c) \quad , \quad b_n = x_n - c$$

آن‌گاه $a_n = o(b_n)$ خواهد بود. یعنی نسبت $\left|\frac{a_n}{b_n}\right|$ به صفر همگرا است و مرتبه‌ی $\{a_n\}$ کوچک‌تر از مرتبه‌ی $\{b_n\}$ است. در رابطه‌ی (۱.۲.۱)، $f(p) = ۰$ است. بنابر این به ازای $x = p$ نتیجه می‌شود:

$$۰ = f(c) + (p-c)f'(c) + o(|p-c|)$$

عبارت $o(|p-c|)$ به دلیل کوچکی، قابل‌چشم‌پوشی است و می‌توان رابطه‌ی زیر را نوشت:

$$۰ \approx f(c) + (p-c)f'(c)$$

با حل رابطه نسبت به p، نتیجه می‌شود:

$$p \approx c - \frac{f(c)}{f'(c)}$$

بنابر این آلگوریتم نیوتون-رافسون به‌صورت زیر طراحی می‌شود:

$$p_{n+۱} = p_n - \frac{f(p_n)}{f'(p_n)} \quad n \geq ۰$$

در گام اول آلگوریتم $p_۰$ به‌عنوان مقدار اولیه‌ی ریشه در نظر گرفته می‌شود، سپس گام‌های بعد تا جایی ادامه می‌یابند که یک دقت کافی نتیجه شود به گونه‌ای که $|p_{n+۱} - p_n| < \varepsilon$ و $\varepsilon > ۰$ باشد. مقدار ε توسط پژوهش‌گر انتخاب می‌شود و به‌طور معمول برای دقت بیش‌تر یک مقدار نزدیک به صفر در نظر گرفته می‌شود.

۳.۲.۱ آمارگیری پساسرشماری

بعد از اتمام سرشماری برای ارزیابی پوشش آن، یک آمارگیری نمونه‌ای از خانوارها با عنوان آمارگیری پساسرشماری انجام می‌شود [بسیاری از کشورها افراد ساکن در اقامتگاه‌های گروهی، منطقه‌های دورافتاده و بی‌خانمان‌ها را از مطالعه حذف می‌کنند]. برای این منظور معمولاً فهرستی مستقل از سرشماری تهیه‌شده (بعضی از کشورها از همان فهرست تهیه‌شده برای اجرای سرشماری استفاده می‌کنند) و از آن نمونه‌ای به اندازه و روش معین انتخاب می‌شود. تمام واحدهای مسکونی که در واحدهای نمونه‌ای (مثلاً بلوک) قرار دارند، P نمونه[۱] را تشکیل می‌دهند. با اجرای آمارگیری پساسرشماری از افراد ساکن در این واحدهای مسکونی بررسی می‌شود که چه افرادی در سرشماری به درستی شمارش شده‌اند، گم‌شده‌اند یا این‌که به‌طور خطاآمیز شمارش شده‌اند. برای این منظور نمونه‌ای از سرشماری لازم است. این نمونه شامل اطلاعات واحدهای مسکونی و افراد ساکن در آن‌ها بوده که در سرشماری شمارش‌شده و به آدرس همان واحدهای نمونه‌ای (مثلاً بلوک) زمین‌کددهی [چه درست و چه غلط] شده‌اند، که E نمونه[۲] را تشکیل می‌دهند. از اطلاعات این دو نمونه در نظام دوگان استفاده می‌شود تا میزان کم‌شماری در سرشماری براورد شود.

۴.۲.۱ اثر خوشه‌بندی

نمونه‌گیری بر اساس واحدهای یک جامعه انجام می‌شود. واحد می‌تواند معادل کوچک‌ترین جزء تشکیل‌دهنده‌ی جامعه یعنی عنصر باشد. اما هنگامی‌که فهرستی از عنصرها در اختیار نیست، نمونه‌گیری بر اساس مجموعه‌ای از عنصرها انجام می‌شود که این نوع واحد نمونه‌گیری را خوشه می‌نامند. یک خوشه

[۱] Population sample (P sample)
[۲] Enumeration sample (E sample)