

شناسایی و کنترل خطاهای بی‌پاسخی و اندازه‌گیری طرح‌های آمارگیری

فرهاد مهران (مجری)

محمد نورمحمدی

زهرا رضایی قهرودی

زهرة فلاح محسن‌خانی

لیدا کلهری ندرآبادی

افسانه یزدانی

محمدرضا ریحانی

تابان باغ‌فلکی

مهدی نقی‌خانی

نریمان یوسفی

روشنک علی‌اکبری صبا



پژوهشکده‌ی آمار

گروه پژوهشی طرح‌های آماری و روش‌های آمار

زمستان ۱۳۹۶

پیش‌گفتار

خطاهای بی‌پاسخی و اندازه‌گیری از مهم‌ترین انواع خطای غیر نمونه‌گیری در آمارگیری‌ها به شمار می‌روند. محاسبه این خطاها و ارائه‌ی راهکارهایی برای کاهش آن‌ها منجر به بهبود نتایج آمارگیری‌ها و حصول برآوردهای کاراتر می‌شود. از آن جایی که محققان و روش‌شناسان آمارگیری همواره به دنبال روش‌های کاراتر برای ارزیابی کیفیت داده‌ها و نتایج حاصل از آمارگیری‌ها هستند، شناسایی منابع مختلف بروز خطاهای بی‌پاسخی و اندازه‌گیری و ارائه‌ی راهکارهایی برای کمی کردن و کاهش این خطاها در آمارگیری‌ها از اهمیت بسیاری برخوردار است.

بر این اساس پژوهشکده‌ی آمار با توجه به رسالت خود در زمینه‌ی اجرای طرح‌های پژوهشی با هدف افزایش کیفیت آمارها، طرح پژوهشی «شناسایی و کنترل خطاهای بی‌پاسخی و اندازه‌گیری طرح‌های آمارگیری» را در دستور کار خود قرار داد. اجرای این طرح پژوهشی از اول تیر ۱۳۹۵ آغاز شده و گزارش نهایی آن اکنون در دسترس علاقه‌مندان قرار گرفته است. این طرح پژوهشی با توجه به سه مورد خاص در ارتباط با آمارگیری نیروی کار مرکز آمار ایران بنا شده است:

۱. اثر خطای بی‌پاسخی و نتایج آزمایش چند روش برای جبران بی‌پاسخی در آمارگیری نیروی کار مرکز آمار ایران،
۲. پرسشگری عین متن و به‌طور کلی استفاده از روش‌های «شناختی» در طراحی پرسشنامه‌ی آماری به‌عنوان روش مؤثر برای یکنواخت کردن خطاهای اندازه‌گیری و نتایج آزمایش این روش در خصوص آمارگیری نیروی کار مرکز آمار ایران،
۳. استفاده از مدل‌های مارکف پنهان به منظور ارزیابی خطاهای اندازه‌گیری در آمارگیری‌های پانلی یا شبه‌پانلی و نتایج عملکرد آن در آمارگیری نیروی کار مرکز آمار ایران.

این پژوهش در گروه پژوهشی طرح‌های فنی و روش‌های آماری پژوهشکده‌ی آمار با همکاری جناب آقای دکتر فرهاد مهران به‌عنوان مجری طرح و آقای دکتر محمد نورمحمدی، خانم دکتر زهرا رضایی قهرودی، خانم دکتر زهره فلاح محسن‌خانی، خانم دکتر لیدا کلهری ندرآبادی، خانم افسانه یزدانی، آقای محمدرضا ریحانی، آقای مهدی نقی‌خانی، آقای نریمان یوسفی، خانم دکتر تابان باغ‌فلکی و خانم دکتر روشنک علی‌اکبری صبا به‌عنوان همکاران اصلی طرح پژوهشی به انجام رسیده است، که بدین وسیله از ایشان صمیمانه تشکر و قدردانی می‌شود.

گروه پژوهشی طرح‌های فنی و روش‌های آمار

پژوهشکده‌ی آمار

فهرست مطالب

مقدمه.....	۱
۱-۱ عملیات آمارگیری.....	۳
۲-۱ مدل سازی ناسازگاری های عملیات آمارگیری.....	۴
۱-۲-۱ مدل گروه های همگن پاسخ.....	۴
۲-۲-۱ مدل اندازه گیری «ساده».....	۵
شناسایی و کنترل خطاهای بی پاسخی.....	۹
۱-۲ منابع ایجاد خطای بی پاسخی و روش های کنترل آن.....	۹
۲-۲ میزان بی پاسخی و روش جبران آن در آمارگیری نیروی کار ایران.....	۱۰
۱-۲-۲ میزان بی پاسخی.....	۱۰
۲-۲-۲ روش جبران بی پاسخی.....	۱۱
۳-۲ روش های متداول برخورد با بی پاسخی در کشورهای دیگر.....	۱۳
۱-۳-۲ روش جایگزینی خانوار بی پاسخ.....	۱۳
۲-۳-۲ روش تعدیل وزن نمونه گیری.....	۱۴
۳-۳-۲ تعدیل بی پاسخی در طرح آمارگیری از نیروی کار در کانادا.....	۱۴
۴-۲ آزمایش چند روش جدید برای طرح آمارگیری نیروی کار در ایران.....	۱۵
۱-۴-۲ انتخاب متغیرها برای ایجاد رده بندی جدید تعدیل بی پاسخی طرح آمارگیری نیروی کار.....	۱۵
۲-۴-۲ شبیه سازی.....	۲۱
۳-۴-۲ نتایج شبیه سازی.....	۲۲
۴-۴-۲ اثرات انجام تعدیل بی پاسخی به روش پیشنهادی در برآوردها.....	۲۵
شناسایی و کنترل خطاهای اندازه گیری.....	۲۹
۱-۳ خطای اندازه گیری.....	۲۹
۱-۱-۳ پرسشنامه.....	۳۰
۲-۱-۳ روش جمع آوری اطلاعات آماری.....	۳۲

۳۳ آمارگیر ۳-۱-۳
۳۴ پاسخگو ۴-۱-۳
۳۴ علوم شناختی ۲-۳
۳۵ کاربرد علوم شناختی در روش‌شناسی آمارگیری ۱-۲-۳
۳۷ آزمایش پرسشگری عین متن در آمارگیری نیروی کار مرکز آمار ایران ۳-۳
۳۸ بازطراحی پرسشنامه طرح نیروی کار ۱-۳-۳
۴۵ روش اجرای پیش‌آزمایش طرح پرسشگری عین متن ۲-۳-۳
۴۵ آموزش در آمارگیری عین متن ۳-۳-۳
۴۶ قدم‌های بعدی ۴-۳
۴۹ برآورد خطاهای اندازه‌گیری ۴-۳
۴۹ طرح آزمایش تصادفی ۱-۴
۵۰ مقایسه با اطلاعات ثبتی ۲-۴
۵۲ مصاحبه مجدد ۳-۴
۵۲ مصاحبه مجدد در آمارگیری نیروی کار مرکز آمار ایران ۱-۳-۴
۵۳ برآورد اریبی پاسخ ۲-۳-۴
۵۳ برآورد واریانس ساده پاسخ ۳-۳-۴
۵۴ مدل مارکوف پنهان ۴-۴
۵۵ مدل‌های مارکوف آشکار ۱-۴-۴
۵۷ مدل‌های کلاس مارکوف پنهان ۲-۴-۴
۶۰ برآورد و آزمون مدل‌های MLC ۳-۴-۴
۶۳ کاربرد مدل MLC در آمارگیری نیروی کار مرکز آمار ایران ۴-۴-۴
۷۱ نتایج و پیشنهادات ۴-۳
۷۱ الف- خطاهای بی‌پاسخی ۴-۳
۷۲ ب- خطاهای اندازه‌گیری ۴-۳
۷۲ ج- برآورد خطاهای اندازه‌گیری ۴-۳
۷۵ پیوست‌ها ۴-۳
۹۹ مرجع‌ها ۴-۳
۱۰۵ واژه‌نامه ۴-۳

فهرست جدول‌ها

- جدول ۱-۲: قدر مطلق میزان تغییرات برآورد نرخ بیکاری با دو روش تعدیل بی‌پاسخی فعلی و پیشنهادی (۳۷۲ رده) برای فصل‌های مختلف سال‌های ۱۳۹۳ تا ۱۳۹۵ ۲۶
- جدول ۲-۲: قدر مطلق میزان تغییرات برآورد نرخ فعالیت اقتصادی با دو روش تعدیل بی‌پاسخی فعلی و پیشنهادی (۳۷۲ رده) برای فصل‌های مختلف سال‌های ۱۳۹۳ تا ۱۳۹۵ ۲۷
- جدول ۱-۳: تقسیم‌بندی خطاهای آماری ۳۰
- جدول ۱-۴: برآورد اثر اریبی پاسخ بر شاخص‌های نیروی کار ۵۳
- جدول ۲-۴: برآورد واریانس ساده پاسخ در آمارگیری نیروی کار ۵۳
- جدول ۳-۴: رده‌بندی متقاطع وضعیت اشتغال در جدول سه طرفه اشتغال (EMP)، بیکاری (UNE) و غیر فعال بودن (NLF) ۵۵
- جدول ۴-۴: تعداد افراد نمونه در سن کار بر حسب وضعیت شغلی مشاهده شده (۱) شاغل، (۲) بیکار، (۳) غیر فعال در چهار دوره چرخش- آمارگیری نیروی کار ۹۲۰۴-۹۳۰۱-۹۳۰۴-۹۳۰۱-۹۴۰۱ ۶۴
- جدول ۵-۴: برآورد خطای اندازه‌گیری نیروی کار بر اساس مدل کلاس مارکف پنهان- مدل (۱) ۶۵
- جدول ۶-۴: برآورد خطای اندازه‌گیری نیروی کار بر اساس مدل کلاس مارکف پنهان- مدل (۲) ۶۶
- جدول ۷-۴: برآورد خطای اندازه‌گیری نیروی کار بر اساس مدل کلاس مارکف پنهان- مدل (۳) ۶۷
- جدول ۸-۴: مقایسه خطای نوع اول و دوم در اشتغال و بیکاری براساس مصاحبه مجدد و مدل مارکوف پنهان (مدل ۳) ۶۸

فهرست شکل ها

- شکل ۱-۱: اساس طرح آمارگیری ۳
- شکل ۲-۱: اساس طرح آمارگیری با خطای اندازه‌گیری ۶
- شکل ۱-۲: نرخ بی‌پاسخی خانوارهای مناطق شهری و روستایی در آمارگیری نیروی کار بر حسب فصل آمارگیری ۱۱
- شکل ۲-۲: نرخ بی‌پاسخی در طرح آمارگیری نیروی کار به تفکیک استان‌های کشور بر اساس نتایج سال‌های ۱۳۹۲ تا ۱۳۹۵ ۱۶
- شکل ۳-۲: نرخ بیکاری در طرح آمارگیری نیروی کار به تفکیک استان‌های کشور بر اساس نتایج سال ۱۳۹۲ تا ۱۳۹۵ ۱۷
- شکل ۴-۲: نرخ بی‌پاسخی در طرح آمارگیری نیروی کار به تفکیک منطقه (شهر مرکز/ سایر مناطق شهری/ مناطق روستایی) در هر فصل ۱۸
- شکل ۵-۲: نرخ بیکاری در طرح آمارگیری نیروی کار به تفکیک منطقه (شهر مرکز/ سایر مناطق شهری/ مناطق روستایی) در هر فصل ۱۸
- شکل ۶-۲: نرخ بی‌پاسخی در طرح آمارگیری نیروی کار به تفکیک نوبت آمارگیری در هر فصل ۲۰
- شکل ۷-۲: نرخ بیکاری در طرح آمارگیری نیروی کار به تفکیک نوبت آمارگیری در هر فصل ۲۰
- شکل ۸-۲: برآورد درصد قدر مطلق اریبی نسبی بی‌پاسخی در طرح آمارگیری نیروی کار برای روش فعلی و پیشنهادی (۳۷۲ رده) تعدیل بی‌پاسخی ۲۲
- شکل ۹-۲: برآورد درصد ریشه میانگین توان‌های دوم خطای نسبی در طرح آمارگیری نیروی کار برای روش فعلی و پیشنهادی (۳۷۲ رده) تعدیل بی‌پاسخی ۲۳
- شکل ۱۰-۲: برآورد درصد قدر مطلق اریبی نسبی بی‌پاسخی در طرح آمارگیری نیروی کار برای روش فعلی و پیشنهادی (۴۱۲ رده) تعدیل بی‌پاسخی ۲۴
- شکل ۱۱-۲: برآورد درصد ریشه میانگین توان‌های دوم خطای نسبی در طرح آمارگیری نیروی کار برای روش فعلی و پیشنهادی (۴۱۲ رده) تعدیل بی‌پاسخی ۲۵
- شکل ۱-۳: چرخه‌ی ساده‌ی آمارگیری ۳۶
- شکل ۲-۳: فرایند ذهنی ۳۶

شکل ۴-۱: نمودار مسیر برای مدل کلاس پنهان مارکوفی..... ۵۹

مقدمه

نظام آمار رسمی ایران مثل سایر کشورهای جهان عمدتاً بر اساس جمع‌آوری اطلاعات از طریق پرسش از افراد در محل اقامت (خانوار) یا در محل کار (کارگاه) یا از طریق منابع ثبتي است. از جمله طرح‌های آمارگیری مستمر از خانوار مرکز آمار ایران، طرح آمارگیری فصلی نیروی کار و طرح آمارگیری سالانه هزینه و درآمد خانوارهای شهری و روستایی است. از جمله طرح‌های آمارگیری از کارگاه، طرح آمارگیری از کارگاه‌های صنعتی دارای ۱۰ نفر کارکن و بیشتر و طرح آمارگیری از معدن در بهره‌برداری کشور است. به‌عنوان مثال آمارگیری از طریق منابع ثبتي طرح پروانه‌های ساختمانی و گزارش فصلی تجارت خارجی را می‌توان نام برد.

از بدو تأسیس، مرکز آمار ایران برای بهبود کیفیت آمارهای رسمی کشور سعی فراوانی داشته و اخیراً هرساله گزارش «بهبود کیفیت آمارها» را در دسترس می‌گذارد. در گزارش سال ۱۳۹۵ فعالیت‌هایی که در جهت تحقق بهبود کیفیت آمارهای رسمی انجام شده است به ترتیب ذیل تشریح گردیده:

- استانداردسازی تعارف واژه‌های اقلام آماری دستگاه‌های اجرایی کشور
- پیاده‌سازی و بازنگری طبقه‌بندی‌های بین‌المللی آماری
- نظارت بر آمارگیری‌ها
- تهیه، تدوین و پیاده‌سازی آزمایشی چارچوب کیفیت آمارهای مرکز آمار ایران
- ارزیابی کیفیت آمارگیری‌های نمونه‌ای
- تدوین گزارش ارزیابی عملکرد مرکز آمار در بعد شاخص‌های عمومی و اختصاصی دستگاه‌های اجرایی کشور.

بدون شک کوشش‌هایی که در استانداردسازی و نظارت و ارزیابی انجام می‌شود، در بهبود کیفیت آمارها مؤثر است. البته خطای آماری در ذات جمع‌آوری اطلاعات آماری است و از بین بردن آن یا کاستن آن مستلزم شناسایی دقیق منابع خطا و کنترل عوامل تأثیرگذار است. اطلاعات جمع‌آوری شده چه از طریق پرسش چه از طریق ثبتي طبیعتاً نمی‌تواند بدون خطا باشد. مثلاً امکان دارد خانوار یا کارگاه تعیین شده برای مصاحبه از دادن

اطلاعات جلوگیری نماید یا فرد واجد شرایط برای پاسخگویی در موقع مصاحبه غایب باشد که در هر دو صورت خطای بی‌پاسخی رخ خواهد یافت، اما چنانچه پاسخ‌گیری در عمل با موفقیت صورت گیرد هم کیفیت پاسخ در همه شرایط یکسان نیست. مثلاً پاسخ می‌تواند به دلائل فراموشی یا ذهنی همراه با خطا باشد. خطا نیز می‌تواند به دلائل درست نفهمیدن سؤال یا در دست نداشتن اطلاعات یا به دلائل عمدی رخ دهد. این نوع خطا در مقابل خطای بی‌پاسخی، خطای پاسخ یا به‌طور کلی‌تر خطای اندازه‌گیری نامیده می‌شود.

خطاهای بی‌پاسخی و اندازه‌گیری دو رده از مهم‌ترین انواع خطا در آمارگیری‌ها هستند که مورد توجه این گزارش هستند. در فصل ۱ موقعیت خطاهای بی‌پاسخی و اندازه‌گیری در کل عملیات آمارگیری نمونه‌ای تشریح می‌گردد و دو نکته اساسی نتیجه‌گیری می‌شود:

- خطای اندازه‌گیری الزاماً با ازدیاد تعداد نمونه از بین نمی‌رود، بلکه در بسیاری از شرایط آریبی و حتی واریانس برآوردها بالا می‌رود.
- کوشش در یکنواخت کردن خطاهای اندازه‌گیری در آمارگیری‌های مستمر نقش مهمی در مهار کردن آریبی و واریانس روندها و تغییرات حاصل از نتایج آماری دارد.

در فصل ۲ این مدل کلی در مورد خطای بی‌پاسخی بررسی و نتایج آزمایشی چند روش جدید برای جبران بی‌پاسخی در آمارگیری نیروی کار تشریح می‌گردد.

در فصل ۳ خطاهای اندازه‌گیری مورد بررسی قرار می‌گیرد. در این فصل نتایج پیش‌آزمایشی «پرسشگری عین‌متنی» به‌عنوان روشی مؤثر برای یکنواخت کردن خطای اندازه‌گیری در آمارگیری نیروی کار و به‌طور کلی استفاده از روش‌های «شناختی» در طراحی پرسشنامه به منظور کنترل خطاهای اندازه‌گیری تشریح می‌شود.

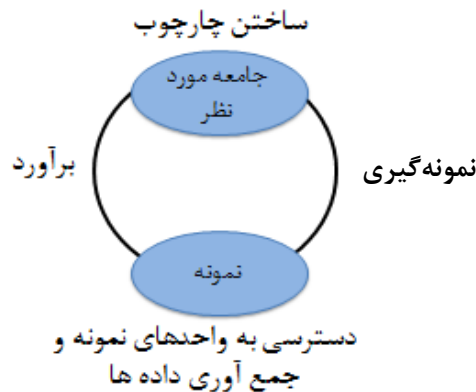
در فصل ۴ روش ارزیابی خطای اندازه‌گیری در آمارگیری‌های پانلی بر اساس مدل‌های مارکوف پنهان معرفی می‌شود و نتایج عمل‌کرد آن در آمارگیری نیروی کار با نمونه چرخشی ارائه می‌گردد. در فصل ۵ بعد از نتیجه‌گیری کلی پیشنهادات اصلی حاصل از این تحقیق بیان می‌شود.

پیشنهاد این طرح توسط خانم دکتر زهرا رضایی قهرودی در پژوهشکده‌ی آمار تهیه گردید و زیر نظر خانم دکتر روشنک علی‌اکبری صبا مدیر گروه طرح‌های فنی و روش‌های آماری با مشاوره آقای دکتر فرهاد مهران و همکاری خانم افسانه یزدانی و آقای محمدرضا ریحانی از مرکز آمار ایران (در بخش خطای بی‌پاسخی)، خانم دکتر زهره فلاح محسن‌خانی و آقای مهدی نقی‌خانی از پژوهشکده‌ی آمار (در بخش پرسشگری عین‌متنی) و خانم دکتر زهرا رضایی قهرودی، خانم دکتر لیدا کلهری ندرآبادی، خانم دکتر زهره فلاح محسن‌خانی و خانم دکتر تابان باغ‌فلکی از پژوهشکده‌ی آمار و دانشگاه تربیت مدرس (در بخش مدل‌های مارکوف پنهان) اجرا شد. پیشرفت کار با برگزاری جلسه‌های کارشناسی متعدد در پژوهشکده‌ی آمار صورت گرفت. گزارش حاضر حاصل این مطالعات و بررسی‌ها است.

۱-۱ عملیات آمارگیری

عملیات یک طرح آمارگیری را می‌توان به چهار دسته تقسیم نمود:

- ۱- ساختن چارچوب جامعه مورد نظر
 - ۲- نمونه‌گیری
 - ۳- دسترسی به واحدهای نمونه و جمع‌آوری اطلاعات
 - ۴- فناوری داده‌ها و محاسبه برآوردها.
- ساختار این عملیات در نمودار ذیل نمایش داده شده است:



شکل ۱-۱: اساس طرح آمارگیری

در آمارگیری با شرایط ایده‌آل فرض می‌شود که:

۱- چارچوب کامل است و رابطه یک به یک با جامعه مورد نظر دارد، یعنی هیچ واحدی از قلم نیافتاده و تمامی واحدهای جامعه مورد نظر در چارچوب بدون تکرار منظور شده‌اند و همچنین هر واحد چارچوب نماینده یک و فقط یک واحد جامعه مورد نظر است.

۲- در نمونه‌گیری هر واحد چارچوب احتمال انتخاب غیر صفر و مشخصی دارد.

۳- در جمع‌آوری اطلاعات دسترسی به تمام واحدهای نمونه میسر است و داده‌های به دست آمده مقادیر واقعی هستند.

۴- فناوری نتایج و محاسبه برآوردها بدون خطا انجام می‌گیرد.

در این شرایط ایده‌آل تنها خطای ممکن خطای نمونه‌گیری است و می‌توان برآورد صفت مورد نظر را بدون اریب و با دقت مشخصی به دست آورد. تنها مقادیر لازم برای این محاسبه احتمال انتخاب π_k و احتمال توأم انتخاب π_{ki} واحدهای نمونه است. به‌عنوان مثال برای یک صفت مورد نظر Y برآورد نااریب مجموع آن در جامعه بر اساس مقادیر به دست آمده از نمونه $y_1, \dots, y_k, \dots, y_n$ به‌صورت ذیل به دست می‌آید:

$$\hat{t} = \sum_{k=1}^n \frac{y_k}{\pi_k}$$

همچنین واریانس این برآورد به صورت ذیل قابل برآورد است:

$$\hat{Var}(\hat{t}) = \sum_{k=1}^n \sum_{l=1}^n \frac{(\pi_{kl} - \pi_k \pi_l)}{\pi_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$$

که در آن $\pi_{kk} = \pi_k$ و n حجم نمونه است.

در عمل در بیشتر آمارگیری‌ها شرایط ایده‌آل همیشه وجود ندارد. اغلب چارچوب ناقص است یا ناسازگاری‌هایی دارد. همه واحدهای چارچوب احتمال انتخاب غیر صفر ندارند و در نمونه‌گیری‌های چند مرحله‌ای تعیین مرزهای واحدهای اولیه با مشکل مواجه می‌شود. در جمع‌آوری اطلاعات به دلایل غیبت موقت و عدم همکاری و غیره برخی از واحدهای نمونه قابل دسترسی نیستند و پاسخ‌های واحدهای مورد دسترسی در موارد گوناگون همراه با خطا هستند. همین طور در فناوری نتایج و ورود داده‌ها و محاسبه برآوردها ناگزیر خطا پیش می‌آید.

۱-۲ مدل‌سازی ناسازگاری‌های عملیات آمارگیری

در این شرایط مسئله‌ای که پیش می‌آید این است که چگونه باید این ناسازگاری‌های عملیاتی را در داخل ساختار آمارگیری که تا به حال فقط خطای نمونه‌گیری را در نظر می‌گرفت، گنجانند؟ در حال حاضر نظریه‌ی کاملی برای برخورد با این مسئله وجود ندارد، اما می‌توان با در نظر گرفتن مدل‌سازی منبع ایجاد خطا برای مورد خاص جواب‌های مفیدی به دست آورد. در این گزارش دو مورد اصلی خطای بی‌پاسخی و خطای اندازه‌گیری مورد بررسی قرار می‌گیرد.

۱-۲-۱ مدل گروه‌های همگن پاسخ^۱ (سارندال و همکاران، ۱۹۹۴)

بر اساس چند فرض ساده می‌توان مدل پاسخ‌گویی مفیدی به نام مدل گروه‌های همگن پاسخ به دست آورد. فرض می‌شود که نمونه به دست آمده $s = \{1, 2, \dots, k, \dots, n\}$ را می‌توان به H_0 گروه $\{h = 1, 2, \dots, H_0\}$ تقسیم نمود، به طوری که احتمال پاسخ‌گویی در داخل گروه یکسان و مابین گروه‌ها مختلف باشد. در ضمن فرض می‌شود که پاسخ‌گویی واحدهای نمونه به طور مستقل از هم صورت می‌گیرد. در این شرایط احتمال پاسخ‌گویی هر واحد نمونه را می‌توان به صورت ذیل بیان کرد.

$$\Pr(k \in r | s) = \pi_{k|s} \quad \forall k \in s$$

$$\Pr(k, l \in r | s) = \pi_{kl|s} = \Pr(k \in r | s) \Pr(l \in r | s) \quad \forall k \neq l \in s$$

که r نشانگر مجموعه واحدهای نمونه که از آنان پاسخ گرفته شده، است. مجموعه پاسخ‌گویان نمونه را می‌توان به گروه‌های همگن تقسیم نمود:

$$r = \bigcup_{h=1}^{H_s} r_h$$

که در آن r_h زیر مجموعه پاسخ‌گویانی هستند که در زیر نمونه همگن s_h قرار دارند. تعداد این واحدهای پاسخ‌گو را با علامت m_h نشان می‌دهیم و

$$m = \sum_{h=1}^{H_0} m_h$$

تعداد کل واحدهای پاسخ‌گو در نمونه است.

اگر مدل صادق باشد، بی‌پاسخی به صورت داده‌های گمشده تصادفی در هر زیر گروه نمونه رخ می‌دهد و توزیع شرطی مجموعه پاسخ‌گویان r به صورت یک طرح نمونه‌گیری با طبقه‌بندی و احتمال برنولی^۲ است، می‌توان احتمالات شرطی را به صورت ذیل نوشت:

$$\Pr(k \in r | s, m) = \pi_{k|s} = \frac{m_h}{n_h} \quad \forall k \in s_h$$

$$\Pr(k, l \in r | s, m) = \pi_{kl|s, m} = \begin{cases} \frac{m_h}{n_h} \frac{(m_h - 1)}{(n_h - 1)} & k \neq l \in s_h \\ \frac{m_h}{n_h} \frac{m'_h}{n'_h} & k \in s_h, l \in s'_h, h \neq h' \end{cases}$$

نسبت‌های $\frac{m_h}{n_h}$ فاکتورهای تعدیل وزن تلقی می‌شوند و برآورد تعداد t در شرایط بی‌پاسخی به صورت ذیل

محاسبه می‌شود:

$$\hat{t} = \sum_{h=1}^{H_s} \sum_{r_h} \frac{n_h}{m_h} \frac{y_k}{\pi_k}$$

در شرایط مدل این برآورد نارایب است و واریانس آن بر اساس فاکتورهای تعدیل وزن قابل محاسبه است. در عمل گروه‌های همگن پاسخ را باید طوری تشکیل داد که با فروض مدل حتی‌الامکان همخوانی داشته باشد. لازم به ذکر است که گروه‌های همگن پاسخ H_s مشروط به ساختار نمونه هستند و می‌توانند از یک نمونه به یک نمونه دیگر تغییر کنند. در فصل (۲) این مطلب در قالب آمارگیری نیروی کار بررسی می‌شود و نتایج آن ارائه می‌گردد.

۱-۲-۲ مدل اندازه‌گیری «ساده»^۳

در آمارگیری علاوه بر مرحله جمع‌آوری اطلاعات، اندازه‌گیری صفات مورد نظر نیز بر اساس ساختار (Process) خاصی صورت می‌گیرد که کنترل آن تماماً در دست طراح نیست. لذا مفید است که مراحل آمارگیری در نمودار (۱) را با شکافتن مرحله جمع‌آوری اطلاعات به صورت نمودار (۲) نمایش دهیم. به این ترتیب یک ساختار (p) وجود دارد که نمونه را ایجاد می‌کند و یک ساختار دیگر (m) داده‌ها را.

^۲ Stratified Bernoulli Sample

^۳ The Simple Measurement Model