دانشگاه علامه طباطبایی

دانشکده‌ی اقتصاد

گروه آمار، ریاضی و علوم کامپیوتر

پایان‌نامه برای دریافت درجه کارشناسی ارشد

در

آمار اجتماعی- اقتصادی

موضوع:

براورد میانگین جامعه در حضور بی‌پاسخی با استفاده از روش نمونه‌گیری دوفازی

پژوهش‌گر: مریم تقربی

استاد راهنما: دکتر نادر نعمت‌الهی

استاد مشاور: دکتر حمیدرضا نواب‌پور

بهمن ماه ۱۳۹۰

به نام خداوند

بخشاینده و مهربان

تایید پایان‌نامه کارشناسی ارشد توسط دانشجو

عنوان پایان‌نامه: براورد میانگین جامعه در حضور بی‌پاسخی با استفاده از روش نمونه‌گیری دوفازی

نام دانشجو: مریم تقربی

شماره‌ی دانشجویی: ۸۸۱۲۵۱۱۶۱۰۴

نام استاد راهنما: دکتر نادر نعمت‌الهی

این‌جانب مریم تقربی دانشجوی کارشناسی ارشد رشته‌ی آمار اجتماعی- اقتصادی دانشکده‌ی **اقتصاد** دانشگاه **علامه طباطبایی** گواهی می‌نمایم پژوهش‌های ارائه 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 ارائه 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 کارشناسی ارشد محروم می‌دانم و هیچ‌گونه ادعایی نخواهم داشت.

تاریخ: امضا دانشجو

کلیه‌ی حقوق مادی و معنوی اعم از چاپ و تکثیر، نسخه برداری، ترجمه،

اقتباس و ... از این پایان‌نامه برای دانشگاه علامه طباطبایی محفوظ است.

نقل مطالب با ذکر منبع مانعی ندارد.

اهدا یه

پیش کش تمام شب‌های بیداری . . .

پیش کش تمام انتظارها . . .

پیش کش دستان ترکیده ی پدرم از . . .

پیش کش دستان لرزان مادرم که . . .

پیش کش به اندیشه های بزرگ استادم . . .

سپاس خدای را که هر توفیقی در گرو عنایت اوست. اکنون که با یاری او توانسته‌ام تلاشی هر چند ناچیز را در راه کسب دانش به انجام رسانم، بر خود لازم می‌دانم از استاد راهنمای بزرگوارم، جناب آقای دکتر نادر نعمت الهی، که به پایان رساندن این تحقیق جز با راهنمایی‌های پدرانه و هدایت‌های بی‌دریغ ایشان میسر نبود، قدردانی نمایم.
از استاد مشاور بزرگوارم، جناب آقای دکتر حمیدرضا نواب‌پور که راهنمایی‌ها و تذکراتشان باعث غنای پایان‌نامه شد، تشکر می‌نمایم.
هم‌چنین از سرکار خانم دکتر روشنک علی‌اکبری صبا و جناب آقای دکتر فرزاد اسکندری که زحمت داوری این اثر را به عهده داشتند سپاس گزارم.
در پایان، از خانواده‌ام، به‌ویژه پدر و مادرم که با حمایت‌های خویش، همواره مرا پشتیبانی کرده‌اند نهایت سپاس و قدرشناسی را دارم.

امیدوارم بتوانم از عهده ادای حق این عزیزان برآیم.

بهمن ۹۰

فهرست مطالب

فهرست مطالب		ب
فهرست جدول‌ها		ت
فهرست شکل‌ها		ج
۱	آشنایی	۱
	۱-۱ مقدمه	۱
	۱-۲ خطاهای نمونه‌گیری و غیر نمونه‌گیری	۳
	۱-۲-۱ انواع خطاهای غیر نمونه‌گیری	۴
	۱-۳ بیان مسئله	۷
	۱-۴ تعریف مفهوم‌ها و واژه‌های اساسی	۱۱
	۱-۵ مرور نوشتگان	۱۱

۱-۶ هدف پژوهش ۱۳

۱-۷ چشم‌انداز فصل‌های آینده ۱۳

۲ براورد میانگین جامع‌ای در حضور بی‌پاسخی با استفاده از اطلاعات یک متغیر کمکی ۱۴

۲-۱ مقدمه ۱۴

۲-۲ روش مراجعه‌ی مجدد ۱۵

۲-۳ نمونه‌گیری دوفازی ۱۹

۲-۴ براوردگرهایی که از اطلاعات یک متغیر کمکی استفاده می‌کنند ۲۲

۲-۵ براوردگرهای پیش‌نهاد شده در حضور بی‌پاسخی در حالت استفاده از اطلاعات یک متغیر کمکی ۲۵

۲-۶ ویژگی براوردگرهای معرفی شده ۲۷

۲-۷ براوردگرهای پیش‌نهاد شده توسط سینگ و کومار (۲۰۰۸) ۳۹

۲-۸ براوردگرهایی بر اساس براورد مقدارهای بهین ۴۵

۲-۹ مقایسه‌ی کارایی براوردگرها ۴۸

۲-۱۰ تعین مقدارهای بهین n, n', k ۵۳

2-11 مثال‌های عددی ۵۹

۲-۱۲ خلاصه‌ی فصل ۶۳

۳ براورد میانگین جامعه‌ای در حضور بی‌پاسخی با استفاده از اطلاعات دو متغیر کمکی ۶۵

۳-۱ مقدمه ۶۵

۳-۲ براوردگرهای پیش‌نهاد شده در حالت استفاده از اطلاعات دو متغیر کمکی در حضور بی‌پاسخی ۶۶

۳-۳ ویژگی براوردگرهای پیش‌نهاد شده ۶۹

۳-۴ مقایسه‌ی کارایی براوردگرها ۸۴

۳-۵ مثال عددی ۹۲

۳-۶ خلاصه‌ی فصل ۹۴

۴ یک کاربرد ۹۵

۴-۱ مقدمه ۹۵

۴-۲ معرفی داده‌ها ۹۵

۴-۳ مطالعه‌ی شبیه‌سازی ۹۶

۴-۳-۱ اندازه‌ی نمونه‌ای، همبستگی و نرخ بی‌پاسخی ۹۸

۴-۳-۲ شبیه‌سازی ۹۹

۴-۴ نتیجه‌گیری ۱۰۱

مرجع‌ها ۱۰۷

واژه‌نامه‌ی فارسی به انگلیسی ۱۱۲

پیوست الف جدول‌ها ۱۱۸

پیوست ب برنامه‌ی شبیه‌سازی ۱۲۴

فهرست جدول‌ها

۲-۱ پارامترهای حاصل از مطالعه‌ی رشد فیزیکی دانش‌آموزان ۶۰

۲-۲ پارامترهای حاصل از مطالعه‌ی تعدادی روستا ۶۰

۲-۳ هزینه‌ی مطالعه‌ی هر واحد نمونه‌گیری در هر فاز (به دلار) ۶۰

۲-۴ درصد کارایی نسبی براوردهای $\overline{Y}_U$ نسبت به $\overline{y}_n^*$ ۶۱

۲-۵ در نظر گرفتن تابع هزینه‌ی ثابت و مقایسه‌ی کارایی براوردها، در نظر گرفتن واریانس ثابت و به‌دست آوردن مقدار هزینه‌ی مورد نیاز برای جامعه‌ی ۱ ۶۳

۲-۶ در نظر گرفتن تابع هزینه‌ی ثابت و مقایسه‌ی کارایی براوردها، در نظر گرفتن واریانس ثابت و به‌دست آوردن مقدار هزینه‌ی مورد نیاز برای جامعه‌ی ۲ ۶۳

۳-۱ پارامترهای حاصل از مطالعه رشد فیزیکی دانش‌آموزان ۹۲

۳-۲ درصد کارایی نسبی براوردگرهای $\overline{Y}_U$ نسبت به $\overline{y}_n^*$ برای مقدارهای متفاوت k ۹۳

۴-۱ SMSE براوردگرها برای همبستگی ۰/۵ و نرخ بی‌پاسخی ۰/۱ ۱۰۳

۴-۲ SMSE براوردگرها برای همبستگی ۰/۵ و نرخ بی‌پاسخی ۰/۲ ۱۰۴

۴-۳ SMSE براوردگرها برای همبستگی ۰/۵ و نرخ بی‌پاسخی ۰/۳ ۱۰۴

۴-۴ SMSE براوردگرها برای همبستگی ۰/۷ و نرخ بی‌پاسخی ۰/۱ ۱۰۴

۴-۵ SMSE براوردگرها برای همبستگی ۰/۷ و نرخ بی‌پاسخی ۰/۲ ۱۰۴

۴-۶ SMSE براوردگرها برای همبستگی ۰/۷ و نرخ بی‌پاسخی ۰/۳ ۱۰۵

۴-۷ SMSE براوردگرها برای همبستگی ۰/۹ و نرخ بی‌پاسخی ۰/۱ ۱۰۵

۴-۸ SMSE براوردگرها برای همبستگی ۰/۹ و نرخ بی‌پاسخی ۰/۲ ۱۰۵

۴-۹ SMSE براوردگرها برای همبستگی ۰/۹ و نرخ بی‌پاسخی ۰/۳ ۱۰۵

۴-۱۰ براوردهای توزیع نمونه‌گیری و توان دوم اریبی آن‌ها برای همبستگی ۰/۵ و نرخ بی‌پاسخی ۰/۲ و $n = 200$ ۱۰۶

۴-۱۰ براوردهای توزیع نمونه‌گیری و توان دوم اریبی آن‌ها برای همبستگی ۰/۵ و نرخ بی‌پاسخی ۰/۲و $n = 500$ ۱۰۶

فهرست شکل‌ها

۱-۱ جامعه‌ی هدف، چارچوب نمونه‌گیری و جامعه‌ی نمونه‌گیری شده ۴

۲-۱ روش مراجعه‌ی مجدد ۱۶

۲-۲ نمونه‌گیری دوفازی با استفاده از اطلاعات یک متغیر کمکی ۲۰

۲-۳ نمونه‌گیری دوفازی با شرط وجود بی‌پاسخی برای متغیر مورد مطالعه‌ی Y و متغیر کمکی X در فاز دوم ۲۱

۲-۴ نمونه‌گیری دوفازی با شرط وجود بی‌پاسخی برای متغیر مورد مطالعه‌ی Y در فاز دوم ۲۲

۲-۵ مقایسه‌ی درصد کارایی براوردهای $\overline{Y}_U$ نسبت به $\overline{y}_n^*$ برای جامعه‌ی (۱) ۶۲

۲-۶ مقایسه‌ی درصد کارایی براوردهای $\overline{Y}_U$ نسبت به $\overline{y}_n^*$ برای جامعه‌ی (۲) ۶۳

چکیده

در آمارگیری‌های نمونه‌ای اطلاعات از برخی واحدها حتی پس از مراجعه‌ی مجدد به‌دست نمی‌آیند و آمارشناس را با نرخی از بی‌پاسخی روبرو می‌کنند. براورد به‌دست آمده از چنین اطلاعات ناکاملی، به‌ویژه هنگامی که ویژگی‌های پاسخ‌گویان از بی‌پاسخ‌ها متفاوت باشند، اریب است. برای برطرف کردن این مشکل و به‌دست آوردن براوردگرهای مناسب، روش‌های مختلفی وجود دارد از جمله تعدیل موزون و جانهی. در عمل، بی‌پاسخی اغلب با تعدیل موزون یا جانهی، جبران می‌شود. این روش‌ها بر فرضیه‌هایی بر ساختار پاسخ‌گویی استوار هستند. هنسن و هورویتس (۱۹۴۶) روش زیرنمونه‌گیری از بی‌پاسخ‌ها و تلاش برای دریافت اطلاعات از آن‌ها را برای تعدیل اریبی بی‌پاسخی پیش‌نهاد کردند. براوردگر پیش‌نهادی آن‌ها ترکیبی از اطلاعات پاسخ‌گویان و زیرنمونه‌ی انتخابی از بی‌پاسخ‌ها است. همان‌طور که می‌دانیم در براورد پارامترهای جامعه مثل میانگین، تعداد کل و نسبت، استفاده از اطلاعات یک یا دو متغیر کمکی باعث افزایش دقت براوردگرها می‌شود. روش‌های براورد نسبتی، ضربی و رگرسیونی و براوردگرهای هم‌گذاشتی آن‌ها نمونه‌های خوبی در این زمینه هستند.

در این پایان‌نامه با استفاده از فن هنسن و هورویتس (۱۹۴۶)، روش نمونه‌گیری دوفازی برای بهبود براورد میانگین جامعه‌ای در حضور بی‌پاسخی با استفاده از اطلاعات یک و دو متغیر کمکی بررسی و براوردگرهایی پیش‌نهاد شده است. با مقایسه‌ی کارایی این براوردگرها، بهترین براوردگر میانگین جامعه‌ای معرفی می‌شود. در حالت استفاده از یک متغیر کمکی براوردگرهای بهینی که حالتی از براوردگرهای نسبتی و رگرسیونی هستند و دارای واریانس یکسان هستند به عنوان بهترین براوردگرها معرفی می‌شوند. در حالت استفاده از دو متغیر کمکی براوردگر بهین هم‌گذاشتی نسبتی و رگرسیونی به عنوان بهترین براوردگر شناخته می‌شود. علاوه بر آن با در نظر گرفتن تابع هزینه‌ی روش دوفازی، در حالت استفاده از یک متغیر کمکی مقدارهای بهینی برای اندازه نمونه‌ای در هر فاز، به‌دست می‌آید و کارا بودن براوردگر معرفی‌شده، با ثابت بودن تابع هزینه یا واریانس براوردگرها نشان داده می‌شود. در نهایت با یک مطالعه‌ی شبیه‌سازی براوردگرهای پیش‌نهادشده در حالت استفاده از یک متغیر کمکی با هم مقایسه می‌شوند و عمل‌کرد بهترین براوردگر نشان داده می‌شود.

واژگان کلیدی

براورد رگرسیونی؛ براورد ضربی؛ براورد میانگین جامعه؛ براورد نسبتی؛ براورد هم‌گذاشتی؛ متغیر کمکی؛ متغیر مورد مطالعه؛ نمونه‌گیری دوفازی.

فصل ۱

آشنایی

۱-۱ مقدمه

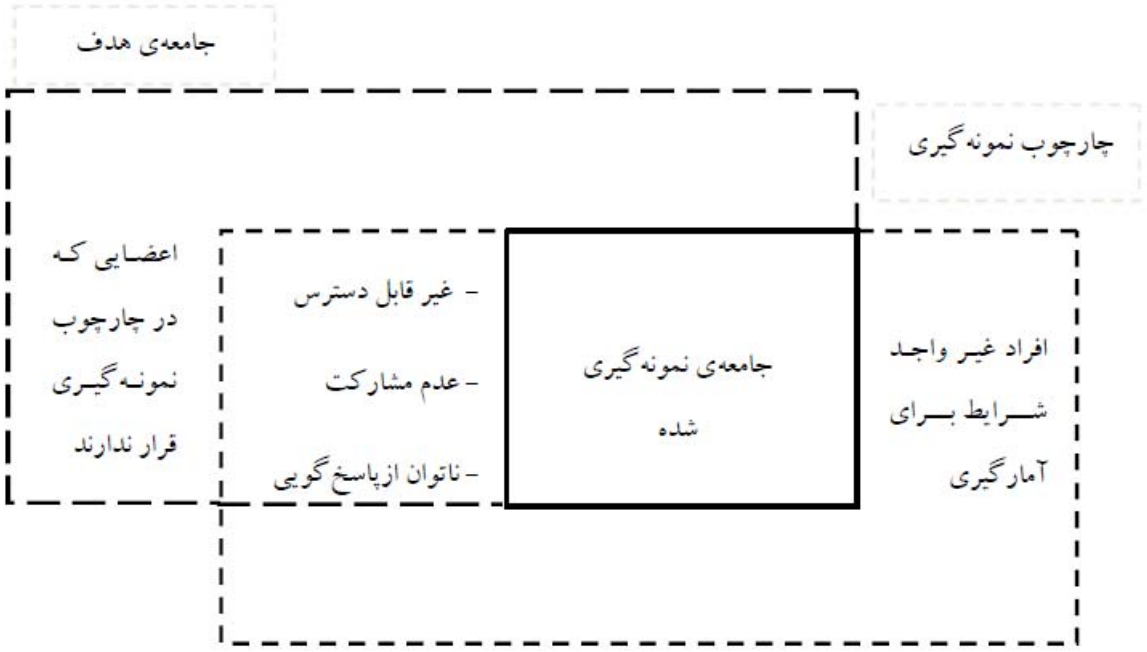
پژوهش‌گران، برنامه‌ریزان، سیاست‌مداران و غیره برای تصمیم‌گیری نیاز به دانستن مشخصه‌هایی (متغیرهایی) از جامعه دارند. برای مثال، درصد بیکاران، تعداد فارغ‌التحصیلان دوره‌ی کارشناسی، میزان تأثیرپذیری دارویی خاص و غیره. برای دست‌یابی به مقدارهای این مشخصه‌ها و انجام استنباط در مورد جامعه نیاز به دانستن اطلاعاتی در مورد آن جامعه است. دست‌یابی به این اطلاعات به شیوه‌های سرشماری، نمونه‌گیری و آمارهای ثبتی صورت می‌پذیرد. در سرشماری از جامعه‌ی متناهی از تمام واحدهای واجد شرایط آمارگیری می‌شود که به هزینه و وقت زیادی نیاز دارد. نمونه‌گیری فرایندی است که در آن بخشی از جامعه انتخاب می‌شود و نتیجه‌های به‌دست آمده از آن را برای کل جامعه تعمیم می‌دهند. اگر نمونه‌گیری پایه و اساس مناسبی برای یک تحقیق باشد، محقق باید اول، جامعه‌ی هدف را برای رسیدن به دقت ممکن تعیین کند و هدف از مطالعه‌ی آن جامعه را روشن سازد. دوم، فهرستی از واحدهای جامعه‌ی هدف تهیه کند تا بتواند نمونه را از آن انتخاب کند. سوم، از یک روش نمونه‌گیری برای انتخاب نمونه‌ی مورد نظر استفاده کند (سالانت و دیلمن، ۱۹۹۴) و چهارم، با استفاده از نتیجه‌های به‌دست آمده از نمونه درباره‌ی جامعه استنباط کند (دِس راج، ۱۹۷۲). تمامی این چهار مرحله همگی باید با هم انجام شوند زیرا نمی‌توان یکی را از دیگری مجزا کرد. نتیجه‌ی فوری که از انجام آن به‌دست می‌آید این است، که نمونه‌ی انتخاب شده باید نماینده‌ی خوبی برای جامعه باشد (سالانت و دیلمن، ۱۹۹۴). در نظر داشته باشید با این‌که اطلاع ما درباره‌ی مشخصه‌های هر جامعه، از گذشته به طور معمول از طریق نمونه‌گیری‌ها فراهم می‌شده است، ولی تا قبل از سال ۱۹۴۰ به مسئله‌ی روش و نظریه‌ی نمونه‌گیری و ارزیابی داده‌های نمونه‌ای توجه چندانی نشده بود. با پیشرفت علم آمار و ارتباط تمام شاخه‌های آمار با نمونه‌گیری، این شاخه از آمار برای براورد مشخصه‌های جامعه و بررسی ویژگی‌های براوردها به سرعت گسترش یافت. افرادی که در گسترش شیوه‌های مختلف نمونه‌گیری به تحقیق پرداخته‌اند در جستجوی راه‌هایی برای انتخاب نمونه‌ای بوده‌اند که اول، نشان‌دهنده‌ی جامعه باشد. دوم، براوردهای مشخصه‌های جامعه‌ای که از آن نمونه گزینش

می‌شود، دقیق باشند و بتوان اعتماد پذیری آن‌ها را اندازه‌گیری کرد و سوم، هزینه‌ی انتخاب این نمونه کم باشد (عمیدی، ۱۳۸۷).

در آمار قاعده‌ی حاکم بر انتخاب نمونه را طرح نمونه‌گیری گویند. طرح نمونه‌گیری عبارت است از انتخاب روشی برای گزینش واحدهای جامعه‌ای برای دست‌یابی به اطلاعات مفید از آن‌ها. طرح‌های نمونه‌گیری را می‌توان بر اساس تعداد مرحله‌های انتخاب واحدها، روش انتخاب واحدها، چگونگی طبقه‌بندی واحدها قبل از انتخاب و موارد دیگر تقسیم‌بندی کرد. طرح‌های نمونه‌گیری پایه‌ای عبارتند از: تصادفی ساده، سیستماتیک و احتمال متناسب با اندازه. در بسیاری از آمارگیری‌ها برای براورد پارامترهای جامعه از نمونه‌گیری تصادفی ساده استفاده می‌شود. حال اگر شناختی از جامعه‌ی مورد مطالعه در دسترس باشد، می‌توان از روش‌های نمونه‌گیری مناسب‌تری برای بالا بردن دقت براوردگرها استفاده کرد. همان‌طور که می‌دانیم آمارگیری‌های نمونه‌ای همیشه با عدم حتمیت همراه هستند زیرا تنها بخشی از جامعه مورد بررسی قرار می‌گیرد و این امر منجر به خطای آمارگیری و اریبی می‌شود. خطای نمونه‌گیری یک مسئله‌ی جدی در کارهای عملی است و آمارشناسان همواره به دنبال کاهش این خطا هستند. یکی از موردهایی که دقت به آن منجر به کاهش این خطا می‌شود، انتخاب مناسب طرح نمونه‌گیری است. نمونه‌گیری احتمالاتی یک طرح نمونه‌گیری خوب است که شامل به‌کارگیری روش‌های احتمالاتی و به حداقل رساندن قضاوت شخصی در انتخاب واحدها برای آمارگیری است. این طرح نمونه‌گیری اریبی را به کمترین مقدار خود می‌رساند که برای آمارشناسان این یک نتیجه مطلوب است.

در یک آمارگیری مطلوب، جامعه‌ی نمونه‌گیری شده مجموعه‌ی تمام واحدهای مشاهده‌ای در انتخاب یک نمونه است که بتواند به وسیله‌ی جامعه‌ی هدف شناسایی شود. جامعه‌ی هدف نیز شامل تمام واحدهایی است که اطلاعاتی از آن‌ها برای متغیرهای مورد نظر گرداوری می‌شود. در آمارگیری از جامعه‌های انسانی، جامعه‌ی نمونه‌گیری شده معمولاً زیرمجموعه‌ی جامعه‌ی هدف است. از آن‌جا که چارچوب نمونه‌گیری فهرستی از تمام واحدهای نمونه‌گیری است، همان‌گونه که در شکل ۱-۱ آشکار است، معمولاً تمام واحدهای جامعه‌ی هدف در چارچوب نمونه‌گیری قرار ندارند، ممکن است تعدادی از واحدها متعلق به چارچوب نمونه‌گیری باشند که در جامعه‌ی هدف قرار نمی‌گیرند و نیز برخی از افراد واجد شرایط در چارچوب نمونه‌گیری، به دلایلی حاضر به پاسخ‌گویی نباشند. برای مثال در آمارگیری تلفنی از افرادی که دارای ویژگی خاصی هستند، چارچوب نمونه‌گیری شامل فهرستی از تمام شماره تلفن‌های موجود است. واضح است که تمام خانوارها دارای تلفن نیستند، بنابر این تعدادی از افراد واجد شرایط در چارچوب قرار نمی‌گیرند. برخی از افراد موجود در چارچوب

دارای ویژگی مورد نظر نیستند و غیر واجد شرایط برای آمارگیری هستند و این‌ها در جامعه‌ی هدف نیستند و برخی از افراد واجد شرایط به دلایلی (عدم مشارکت، عدم دسترسی، بیماری و غیره) مایل به پاسخ‌گویی نیستند (لوهر، ۲۰۱۰). هنگامی که نمونه‌ای از چارچوب انتخاب می‌شود و ما بخواهیم توسط براوردگرهای به‌دست آمده از آن در مورد جامعه نتیجه‌گیری کنیم، به دلیل‌های گفته شده در بالا و دیگر موردها که در ادامه به طور مفصل مورد بررسی قرار خواهند گرفت، حتماً نتیجه‌گیری‌ها با خطا همراه خواهند بود. این خطاها دو نوع هستند، خطای نمونه‌گیری و خطای غیرنمونه‌گیری که در بخش بعد به بررسی این خطاها پرداخته می‌شود.



شکل ۱-۱: جامعه‌ی هدف، چارچوب نمونه‌گیری و جامعه‌ی نمونه‌گیری شده

۱-۲ خطاهای نمونه‌گیری و غیر نمونه‌گیری

در آمارگیری‌ها به طور معمول دو نوع خطا روی می‌دهد: خطای نمونه‌گیری و خطای غیرنمونه‌گیری. درصورتی که سرشماری انجام شود، این آمارگیری دارای خطای نمونه‌گیری نیست اما براوردهای حاصل از آن دارای خطای غیرنمونه‌گیری بالایی هستند. در نمونه‌گیری‌ها هر دو خطا رخ می‌دهند و هر چه تعداد نمونه کمتر باشد، معمولاً خطای غیرنمونه‌گیری کمتری روی می‌دهد.

خطای نمونه‌گیری: این خطا در اثر استنباط در مورد تعداد زیادی از واحدها با استفاده از تعداد اندکی از آن‌ها به وجود می‌آید. در واقع خطای نمونه‌گیری تفاوت بین براورد حاصل از آمارگیری نمونه‌ای با مقدار واقعی پارامتر است.

خطای غیرنمونه‌گیری: به هر خطایی غیر از خطای نمونه‌گیری خطای غیر نمونه‌گیری گفته می‌شود. خطاهای غیرنمونه‌گیری بیش‌تر به شیوه‌ی گرداوری و پردازش داده‌ها وابسته هستند (باندا، ۲۰۰۳). این خطاها به علت تعریف‌ها و مفهوم‌های گمراه‌کننده‌ی پرسش‌ها، چارچوب‌های ناقص، روش‌های نادرست گرداوری داده‌ها، جدول‌بندی، کدگذاری، پوشش ناقص واحدهای نمونه‌ای وغیره رخ می‌دهند. خطای غیرنمونه‌گیری غیرقابل پیش‌بینی است و به سادگی کنترل نمی‌شود و برخلاف خطای نمونه‌گیری ممکن است با افزایش نمونه، افزایش یابد.

۱-۲-۱ انواع خطاهای غیرنمونه‌گیری

در این قسمت برای آشنایی بیش‌تر با خطاهای غیر نمونه‌گیری آن‌ها را معرفی و به صورت مختصر بررسی می‌کنیم. مهم‌ترین خطاهای غیر نمونه‌گیری عبارتند از

(الف) خطای تشخیص

این خطا شامل خطاهای مفهومی است. برای مثال خطا در مفهوم پرسش، طراحی نامناسب پرسش‌نامه، شیوه‌ی نادرست مصاحبه و غیره. خطای تشخیص هنگامی رخ می‌دهد که مفهومی توسط پرسش رسانده شود که با برداشت اصلی که فرد پاسخ‌گو از آن باید داشته باشد، متفاوت باشد.

(ب) خطای پوشانش یا چارچوب

این خطا در اثر اختلاف بین جامعه‌ی هدف و چارچوب نمونه‌گیری به وجود می‌آید و شامل خطای بیش‌پوشانشی و خطای کم‌پوشانشی است. خطای بیش‌پوشانشی به علت‌های زیادی از جمله تکراری بودن واحدهای واجد شرایط در چارچوب نمونه‌گیری رخ می‌دهد. در آمارگیری تلفنی از خانوارهای ساکن یک شهر ممکن است خانواری دارای بیش‌تر از یک خط تلفن باشد و بیش‌تر از یک‌بار برای آمارگیری با او تماس گرفته شود بنابر این نتیجه‌های آمارگیری دارای خطای بیش‌پوشانشی هستند. خطای کم‌پوشانشی زمانی رخ می‌دهد که برخی از واحدهای واجد شرایط و متعلق به جامعه‌ی هدف در چارچوب نمونه‌گیری قرار نگیرند. گاهی اوقات ممکن است چارچوبی برای جامعه‌ی مورد علاقه وجود نداشته باشد، چارچوب ناقص باشد یا چارچوب نمونه‌گیری قدیمی و تاریخ گذشته باشد. تمام این موردها موجب می‌شود که نتوان تمام افراد واجد

شرایط جامعه را به درستی شناسایی کرد بنابر این نتیجه‌های آمارگیری دارای خطای کم‌پوشانشی می‌شوند. آمارشناسان برای کاهش این خطا به‌کارگیری بازه‌ی نیم‌باز، چارچوب‌های چندگانه، تعدیل‌های پساآمارگیری و براوردگرهای کالبیده را پیش‌نهاد کرده‌اند.

(پ) خطای اندازه‌گیری

خطای اندازه‌گیری شامل تمام خطاهایی است که موجب شوند مقدار اندازه‌گیری برای متغیر مورد مطالعه برای یک واحد نمونه‌ای با مقدار واقعی آن متفاوت باشد. خطای اندازه‌گیری شامل خطاهای مرحله‌ی انتخاب نمونه و گرداوری داده‌ها است. انتخاب یک نمونه‌ی سهل و ساده که شامل افرادی است که با احتمال زیاد پاسخ‌گو خواهند بود، انتخاب عمدی یک نمونه‌ی نماینده، تعیین اشتباه جامعه‌ی هدف، جای‌گزینی افرادی که به راحتی از یک جامعه به‌دست می‌آیند با افرادی که دسترسی به آن‌ها آسان نیست، انتخاب نمونه‌ای از داوطلبان برای یک آمارگیری و غیره همگی دارای اریبی انتخاب هستند. بنابر این آمارشناسان برای کاهش این خطا همواره به دنبال نمونه‌ای هستند که نماینده‌ی خوبی برای جامعه باشد. نمونه‌های احتمالاتی دارای مینیمم اریبی انتخاب هستند اما استفاده از آن‌ها بسیار پرهزینه است.

گاهی اوقات هنگام گرداوری داده‌ها از جامعه‌های انسانی، مردم واقعیت را پنهان می‌کنند. برای مثال در آمارگیری‌های کشاورزی، کشاورزان یک ناحیه ممکن است محصولات را کم گزارش کنند و این منجر به اشتباه در اندازه‌گیری متوسط محصولات کشاورزی یک منطقه شود. اگر افراد در مورد سؤالی که از آن‌ها پرسیده می‌شود آگاهی نداشته باشند، ممکن است پاسخ نامناسب بدهند. همچنین جمله‌بندی و تعیین واژه‌های درست، ترتیب مناسب قرار گرفتن سؤال‌ها؛ طراحی مناسب و ظاهر مصاحبه‌گران تأثیر زیادی در دست‌یابی به پاسخ صحیح از افراد دارد؛ زیرا مردم پاسخ‌های متفاوتی به مصاحبه‌گران متفاوت می‌دهند. در واقع هرچه طبقه‌ی اجتماعی بین پاسخ‌گو و مصاحبه‌گر بیش‌تر باشد احتمال ارائه‌ی پاسخ صحیح و کامل به سؤال‌های حساس کمتر است. مردم ممکن است برای پرهیز از خطرهایی که آن‌ها را تهدید خواهد کرد از بازگو کردن نظر واقعی‌شان اجتناب کنند پس هنگام آمارگیری شرایطی باید فراهم شود که موجب اعتماد واحد پاسخ‌گو شود. بهتر است از سؤال‌های بازپاسخ در پرسش‌نامه پرهیز شود زیرا پاسخ‌ها ممکن است ناخوانا و بی‌ربط باشند. همچنین امکان وجود غلط املایی بر مشکلات طبقه‌بندی و تفسیر پاسخ‌های باز می‌افزاید که آن منجر به خطای اندازه‌گیری می‌شود. خطای اندازه‌گیری یک نگرانی برای تمام آمارشناسان است و می‌تواند خیلی مشکل‌ساز شود.

(ت) خطای پردازش

خطای پردازش هنگام برنامه‌نویسی، ورود داده، کدگذاری، ویرایش داده‌ها، جدول‌بندی و غیره رخ می‌دهد. استفاده از سؤال‌های بازپاسخ در پرسش‌نامه نیز ممکن است منجر به خطای پردازش شود برای مثال در دسته‌بندی وتفسیر طبقه‌ی افراد با پرسش سؤال‌های باز امکان دارد دچار اشتباه شویم. هنگام پردازش داده‌ها اگر برای تعدیل بی‌پاسخ‌ها روش‌های مناسبی به‌کار برده نشود، منجر به خطای پردازش می‌شود. برای کاهش این خطا قبل از تجزیه و تحلیل داده‌ها توسط نرم‌افزار باید برنامه را آزمون کرد، داده‌های وارد شده را بازنگری و سؤال‌ها را در یک مقیاس مناسب کدگذاری کنیم.

(ث) خطای بی‌پاسخی

این خطا به علت ناکامی در دست‌یابی به اطلاعات مفید در مورد یک یا چند متغیر مورد بررسی برای یک یا چند عضو واجد شرایط در آمارگیری رخ می‌دهد (بیداربخت‌نیا، ۱۳۸۵). حتی اگر از طرح‌های آمارگیری دقیقی برای مینیمم کردن دیگر منابع خطا استفاده شود، وجود بی‌پاسخی ممکن است تمام نتیجه‌های آمارگیری را خراب کند. به طور کلی افرادی که مصاحبه را ناتمام می‌گذارند و یا به هیچ وجه حاضر به پاسخ‌گویی نیستند متفاوت از افرادی هستند که مصاحبه را تمام می‌کنند. اما وسعت این تفاوت‌ها ناشناخته است مگر آن‌که بتوان به شیوه‌ای اطلاعاتی در مورد بی‌پاسخ‌ها به‌دست آورد. در بیش‌تر موردها، بی‌پاسخی در میان واحدهای نمونه‌ای به طور یکنواخت پراکنده نیست ولی در میان زیرگروه‌ها به شدت متمرکز است. به دلیل اختلاف بین پاسخ‌گویان و پاسخ‌نداده‌ها، توزیع نمونه‌هایی که مقدارهای آن‌ها در میان زیرگروه‌ها اندازه‌گیری شده‌اند از توزیع نمونه‌های انتخاب شده منحرف خواهد شد. همچنین اگر متغیرهای آمارگیری به زیرگروه‌ها وابسته باشند به احتمال زیاد این انحراف اریبی حاصل از بی‌پاسخی را بالا می‌برد (گنجعلی و همکاران، ۱۳۸۶). نرخ بی‌پاسخی برای یک آمارگیری عبارت است از تعداد واحدهای بی‌پاسخ به کل افراد نمونه و نرخ پاسخ‌گویی عبارت است از تعداد واحدهای پاسخ‌گو به کل واحدهای نمونه. بسیاری از آمارگیری‌ها دارای نرخ پاسخ‌گویی کمی هستند برای مثال اگر نرخ پاسخ‌گویی زیر ۱۰٪ باشد بسیار سخت می‌توان نتیجه‌ها را به جامعه‌ای که به ۹۰٪ جامعه‌ی نمونه‌گیری شده نمی‌توان دست یافت؛ تعمیم داد. در بسیاری از مطالعه‌ها برای آن‌که بتوان با به‌کارگیری روش‌های تعدیل، نرخ بی‌پاسخی را تا حد امکان کاهش داد پاسخ‌گویان و بی‌پاسخ‌ها را در دو گروه متفاوت جای می‌دهند یک گروه شامل افراد پاسخ‌گو و گروه دیگر شامل افراد بی‌پاسخ است. پژوهش‌گران با خطای

بی‌پاسخی بیش‌تر در پرسش‌نامه‌های پستی، پس از آن در مصاحبه‌های تلفنی و سپس در مصاحبه‌های فردی روبرو می‌شوند.

3-1 بیان مسئله

همان‌طور که بیان شد، یکی از خطاهای غیرنمونه‌گیری، خطای بی‌پاسخی است که به یکی از علت‌های زیر رخ می‌دهد

۱- دسترسی به واحد نمونه‌گیری واجد شرایط ممکن نباشد،

۲- واحد نمونه‌گیری به دلیل‌هایی حاضر به همکاری نباشد. (در این مورد همواره نمی‌توان واجد شرایط بودن آن واحد را تشخیص داد)،

۳- واحد نمونه‌گیری به دلیل‌های خاصی قادر به پاسخ‌گویی نباشد (حساس بودن پرسش‌ها، قابل فهم نبودن پرسش‌ها، ناتوانی پاسخ‌گو و ...) و

۴- برگشت‌ندادن پرسش‌نامه در آمارگیری به روش پستی.

موارد بالا در دو عامل بی‌پاسخی واحد اطلاعاتی و بی‌پاسخی قلم اطلاعاتی خلاصه می‌شوند. بی‌پاسخی واحد زمانی به وجود می‌آید که هیچ اطلاعی از واحد نمونه‌ای در دست نباشد. برای مثال عدم دسترسی به واحد نمونه‌ای یا امتناع از مشارکت در آمارگیری توسط واحد نمونه‌ای. اگر واحد نمونه‌ای به برخی از سؤال‌های پرسش‌نامه پاسخ دهد و برخی از پرسش‌ها را بی‌پاسخ بگذارد، آمارشناس را با بی‌پاسخی قلم روبرو کرده است. بی‌پاسخی قلم، زمانی که مفهوم سؤالی برای واحد نمونه‌ای روشن نباشد، زمان آمارگیری مناسب نباشد یا فرد فرصت پاسخ به تمام پرسش‌ها را نداشته باشد و موردهایی از این قبیل، رخ می‌دهد. در بی‌پاسخی واحد نسبت به بی‌پاسخی قلم، ما اطلاعات کمتری در مورد واحد نمونه‌ای داریم. رفتار مصاحبه‌گران نیز تأثیر بسیاری در افزایش نرخ پاسخ‌گویی واحد نمونه‌ای دارد. در واقع در آمارگیری هرچه نرخ بی‌پاسخی (قلم یا واحد) پایین‌تر باشد نشان‌دهنده‌ی پذیرش عمومی موضوع آمارگیری توسط واحد نمونه‌ای، سازمان‌دهی و مدیریت خوب آمارشناس، روشن و قابل فهم بودن پرسش‌ها و عملکرد خوب آمارگیران در برخورد با واحد نمونه‌ای، انتخاب زمان مناسب آمارگیری، فن‌های نگارش صحیح پرسش‌نامه و غیره است.

وجود بی‌پاسخی موجب کوچک شدن اندازه‌ی نمونه‌ای و کاهش دقت و اریبی براوردگرها می‌شود. اگر آمارشناسان با رعایت تمام موردها در طرح آمارگیری، به دلیل‌هایی با بی‌پاسخی روبرو شوند با استفاده از تعدیل‌های پساآمارگیری به دنبال حداقل کردن خطای بی‌پاسخی براوردگرها می‌باشند. روش‌های تعدیل بی‌پاسخی عبارت‌اند از: (الف) وزن‌دهی، (ب) جانهی و (ج) مراجعه‌ی مجدد.

(الف) وزن‌دهی: وزن‌دهی وقتی به‌کار می‌آید که پژوهش‌گر با بی‌پاسخی واحد اطلاعاتی روبرو شده باشد و یا تمام متغیرهای مهم برای یک واحد بی‌پاسخ باشند. در این روش وزن واحدهای بی‌پاسخ به واحدهای پاسخ‌گو داده می‌شود. روش پساطبقه‌بندی نیز یکی از روش‌های وزن‌دهی، است.

(ب) جانهی: در آمارگیری‌هایی که بی‌پاسخی قلم آماری دارند، برای تعدیل اثر بی‌پاسخی از جانهی استفاده می‌کنند. در جانهی مقدارهای بی‌پاسخ با مقدارهای قابل قبولی برای به‌دست آوردن یک مجموعه‌ی کامل از داده‌ها به شیوه‌هایی جای‌گزین می‌شوند (بخشنده، ۱۳۸۸). روش‌های جانهی عبارت‌اند از: جانهی استقرایی، جانهی میانگین سراسری، جانهی میانگین طبقه‌ای، جانهی رگرسیونی، جانهی بی‌درنگ، جانهی بادرنگ و جانهی چندگانه. برای مطالعه‌ی جزییات بیش‌تر پیرامون دو روش وزن‌دهی و جانهی به لوهر (۲۰۱۰) مراجعه شود.

(ج) کاهش اثر بی‌پاسخی به روش مراجعه‌ی مجدد: این روش موضوع مورد مطالعه‌ی پژوهش‌گر است. در این روش زیرنمونه‌ای از بی‌پاسخ‌ها انتخاب شده، و براوردهای نهایی با استفاده از داده‌های مربوط به واحدهای پاسخ‌گو و زیرنمونه‌ی انتخابی از بی‌پاسخ‌ها به‌دست می‌آیند. روش مراجعه‌ی مجدد اولین بار توسط هنسن و هورویتس در سال ۱۹۴۶ پیش‌نهاد شد. این روش برای هر طرح نمونه‌گیری قابل استفاده است اما در این پایان‌نامه این روش را برای طرح نمونه‌گیری تصادفی‌ساده بدون جای‌گذاری مورد بررسی قرار می‌دهیم. روش مراجعه‌ی مجدد به صورت زیر استفاده می‌شود.

فرض کنید اندازه‌ی جامعه‌ی متناهی N باشد، از این جامعه نمونه‌ای به اندازه‌ی n به روش تصادفی‌ساده انتخاب و از آن آمارگیری کنیم. فرض کنید با n_1 واحد پاسخ‌گو و n_2 واحد بی‌پاسخ روبرو شویم. براوردگرهای معمول پارامترهای جامعه وقتی بی‌پاسخی وجود دارد، احتمالاً اریب خواهند بود. هنسن و هورویتس (۱۹۴۶) برای کاهش اریبی، پیش‌نهاد کردند که زیرنمونه‌ای به اندازه‌ی n_{2m}؛ که $k>1$,$n_{2m}=\frac{n_2}{k}$ از n_2 واحد بی‌پاسخ انتخاب کرده و به آن‌ها برای دریافت پاسخ مراجعه شود. با فرض این‌که از تمامی n_{2m} واحد انتخابی پاسخ دریافت شود، براوردگر تعدیل‌شده‌ی مجموع جامعه‌ای برای متغیر مورد بررسی Y ، توسط هنسن و هورویتس به صورت زیر ارائه شده است

$$\hat{t}=N\,\hat{\bar{Y}}_U=\frac{N}{n}\left(n_1\bar{y}_1+n_2\bar{y}_{2m}\right) \qquad (1\text{-}3\text{-}1)$$

که در آن

$\bar{y}_1$: میانگین متغیر مورد بررسی برای واحدهایی که در مراجعه‌ی اول پاسخ می‌دهند،

$\bar{y}_{2m}$: میانگین متغیر مورد بررسی برای واحدهایی که در مراجعه‌ی اول پاسخ نمی‌دهند، اما با مراجعه‌ی مجدد پاسخ می‌دهند،

n_1: تعداد واحدهایی که در اولین مراجعه پاسخ می‌دهند و

n_2: تعداد واحدهایی که در اولین مراجعه پاسخ‌گو نیستند اما با مراجعه‌ی مجدد پاسخ‌گو خواهند بود.

هنسن و هورویتس (۱۹۴۶) نشان دادند که براوردگر ۱-۳-۱ برای مجموع جامعه‌ای نااریب است، اما واریانس آن بیش‌تر از واریانس معمولی براوردگر مجموع جامعه‌ای در روش نمونه‌گیری تصادفی ساده است.

هر طرح آماری دارای پیش‌فرض‌هایی است که قبل از اجرای آمارگیری باید در نظر گرفته شوند. پیش‌فرض‌های روش مراجعه مجدد عبارت‌اند از:

۱- توزیع متغیر مورد مطالعه‌ی Y برای واحدهای پاسخ‌گو و بی‌پاسخ یکسان است و

۲- تمام n_{2m} واحد در مراجعه‌ی مجدد به آمارگیری پاسخ می‌دهند.

در تمام روش‌های تعدیل که در بالا معرفی شدند، اگر از متغیرهای کمکی که با متغیر مورد مطالعه همبستگی زیادی دارند، اطلاعاتی در دست باشد می‌توان براوردگرهایی با دقتی بسیار بالاتر به‌دست آورد.

یکی از روش‌های نمونه‌گیری که در آن از روش مراجعه‌ی مجدد برای تعدیل بی‌پاسخی استفاده می‌شود، روش نمونه‌گیری دوفازی است. روش نمونه‌گیری دوفازی در حضور بی‌پاسخی برای متغیر مورد مطالعه، بر اساس انتخاب نمونه‌ای کوچک‌تر از درون نمونه‌ای بزرگ‌تر از جامعه است. هنگامی که برای متغیر مورد مطالعه بی‌پاسخی وجود دارد و اطلاعاتی برای متغیر کمکی در دسترس نیست، می‌توان از ترکیب دو روش دوفازی و مراجعه‌ی مجدد به صورت زیر استفاده کرد و دقت براوردگرها را بهبود بخشید.

فرض کنید Y متغیر مورد مطالعه و X متغیر کمکی باشد. در غیاب اطلاعات کمکی در فاز اول نمونه‌ای نسبتاً بزرگ به اندازه‌ی n' از جامعه انتخاب کرده و اطلاعاتی برای متغیر کمکی X جمع‌آوری کرده و سپس در فاز دوم از آن زیرنمونه‌ای کوچک‌تر به اندازه‌ی n ($n < n'$) انتخاب کرده و اطلاعات را برای متغیر مورد مطالعه ی Y جمع‌آوری می‌کنند، به علت وجود بی‌پاسخی برای متغیر مورد مطالعه از روش مراجعه‌ی مجدد برای تعدیل بی‌پاسخی استفاده می‌شود. فرض کنید با n_1 واحد پاسخ‌گو و n_2 واحد بی‌پاسخ مواجهه می‌شویم. از n_2 واحد بی‌پاسخ، زیرنمونه‌ای به اندازه‌ی n_{2m} انتخاب کرده و به آن‌ها برای کسب اطلاعات بیش‌تر برای متغیر اصلی Y مراجعه می‌شود. انتخاب نمونه‌ای بزرگ از جامعه را فاز اول و انتخاب نمونه‌ای کوچک‌تر از درون نمونه‌ی بزرگ را فاز دوم می‌نامند. با استفاده از روش نمونه‌گیری دوفازی و استفاده از اطلاعات متغیرهای کمکی که با متغیر مورد مطالعه همبستگی بالایی دارند می‌توان دقت براوردگرهای