دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

فاصله اطمینان ناپارامتری در طرح‌های نمونه‌گیری سازوار

رساله دکتری آمار کاربردی

محمد محمدی

استاد راهنما

دکتر محمد صالحی مرزیجرانی

۱۳۹۰

دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

رساله دکتری آمار کاربردی آقای محمد محمدی

تحت عنوان

فاصله اطمینان ناپارامتری در طرح‌های نمونه‌گیری سازوار

در تاریخ ۱۳۹۰/۴/۱۳ توسط کمیته تخصصی زیر مورد بررسی و تصویب نهائی قرار گرفت.

۱– استاد راهنمای پایان‌نامه — دکتر محمد صالحی مرزیجرانی

۲– استاد مشاور پایان‌نامه — دکتر هوشنگ طالبی

۳– استاد داور ۱ — دکتر نادر نعمت‌اللهی (دانشگاه علامه طباطبایی)

()

۴– استاد داور ۲ — دکتر علیرضا نعمت‌اللهی (دانشگاه شیراز)

. ۵– استاد داور داخلی — دکتر سعید پولادساز

سرپرست تحصیلات تکمیلی دانشکده دکتر اعظم اعتماد

تقدیم به:

. زورق‌نشین دریای بی‌کران معرفت

. حضرت قمر بنی هاشم (ع)

کلیه حقوق مادی مترتب بر نتایج مطالعات، ابتکارات و نوآوری‌های ناشی از
تحقیق موضوع این پایان‌نامه متعلق به دانشگاه صنعتی اصفهان است.

فهرست مطالب

فصل اول مقدمه ۲

۱–۱ تعریف‌ها و نمادها ۷

۱–۲ استنباط آماری در جامعه متناهی ۱۰

۱–۳ انواع برآوردگرهای نااریب ۱۵

۱–۳–۱ برآوردگر π (برآوردگر هارویتز–تامپسون) ۱۵

۱–۳–۲ برآوردگر هایک ۱۶

۱–۳–۳ برآوردگر مورتی ۱۷

۱–۳–۴ برآوردگر pwr (برآوردگر هنسن–هرویتز) ۱۸

۱–۴ انواع طرح‌های نمونه‌گیری ۱۹

۱–۴–۱ طبقه‌بندی پسین ۲۳

فصل دوم برآوردهای فاصله‌ای در نمونه‌گیری پیمایشی ۲۶

۲–۱ مقدمه ۲۶

۲–۲ فاصله اطمینان تقریب نرمال ۲۹

۲–۳ روش خطی‌سازی تیلر ۳۲

۲–۴ روش زیرنمونه‌گیری بوت‌استرپ ۳۵

۲–۴–۱ روش بوت‌استرپ در نمونه‌گیری طبقه‌ای ۳۶

۲–۴–۲ روش بوت‌استرپ در نمونه‌گیری با احتمال‌های نابرابر ۴۱

۲–۴–۳ تعیین تعداد تکرار زیرنمونه‌گیری ۴۳

۲–۵ انواع فاصله اطمینان بوت‌استرپ ۴۳

۲–۶ درستنمایی تجربی ۴۶

۲–۷ درستنمایی تجربی در جامعه متناهی ۴۸
۲–۷–۱ نمونه‌گیری تصادفی غیر طبقه‌ای ۴۹
۲–۷–۲ نمونه‌گیری تصادفی طبقه‌ای ۵۱
۲–۸ فاصله اطمینان درستنمایی تجربی ۵۲

فصل سوم طرح‌های نمونه‌گیری معکوس ۵۶
۳–۱ مقدمه ۵۶
۳–۲ طرح‌های نمونه‌گیری معکوس با احتمال برابر ۵۸
۳–۳ نمونه‌گیری معکوس معمولی ۵۹
۳–۴ قوانین توقف متغیر ۶۰
۳–۵ نمونه‌گیری معکوس تعمیم‌یافته ۶۱
۳–۶ برآوردگر مورتی تحت طرح‌های نمونه‌گیری معکوس ۶۱
۳–۶–۱ نمونه‌گیری معکوس معمولی ۶۲
۳–۶–۲ نمونه‌گیری معکوس تعمیم‌یافته ۶۵
۳–۷ برآوردگر هارویتز–تامپسون تحت طرح‌های نمونه‌گیری معکوس ۶۶
۳–۷–۱ نمونه‌گیری معکوس معمولی ۶۶
۳–۷–۲ قوانین توقف متغیر ۶۸
۳–۸ برآوردگر طبقه‌بندی پسین ۷۱
۳–۹ نمونه‌گیری معکوس با جایگذاری ۷۴
۳–۹–۱ برآوردگر مورتی ۷۵
۳–۱۰ مقایسه برآوردگرهای مورتی، هارویتز–تامپسون و طبقه‌بندی پسین ۷۷
۳–۱۱ کارآیی طرح نمونه‌گیری معکوس ۷۹
۳–۱۲ مطالعه شبیه‌سازی ۸۲

فصل چهارم نمونه‌گیری خوشه‌ای سازوار ۹۲
۴–۱ مقدمه ۹۲
۴–۲ معرفی نمونه‌گیری خوشه‌ای سازوار ۹۳
۴–۳ برآوردگرهای نااریب در نمونه‌گیری خوشه‌ای سازوار ۹۶
۴–۳–۱ میانگین نمونه اولیه ۹۷

۴–۳–۲ برآوردگر اصلاح‌شده هنسن–هرویتز ۹۷
۴–۳–۳ برآوردگر اصلاح‌شده هارویتز–تامپسون ۹۸
۴–۳–۴ برآوردگر اصلاح‌شده هایک ۹۹
۴–۴ مقایسه برآوردگرهای هارویتز–تامپسون و هنسن–هرویتز ۱۰۰
۴–۴–۱ توزیع حدی برآوردگرها ۱۰۲
۴–۵ کارآیی نمونه‌گیری خوشه‌ای سازوار ۱۰۴
۴–۶ مقایسه طرح نمونه‌گیری خوشه‌ای سازوار و طرح‌های سنتی ۱۰۸
۴–۷ مطالعه شبیه‌سازی ۱۰۹

فصل پنجم فاصله‌های اطمینان ناپارامتری در نمونه‌گیری معکوس ۱۱۳
۵–۱ مقدمه ۱۱۳
۵–۲ روش بوت‌استرپ در نمونه‌گیری معکوس ۱۱۴
۵–۳ بوت‌استرپ نمونه معکوس در حالت M معلوم: برآوردگر هارویتز–تامپسون ۱۱۵
۵–۳–۱ روش بوت‌استرپ با جایگذاری (BWR) ۱۱۶
۵–۳–۲ روش بوت‌استرپ آیینه–انطباق سیتر ۱۱۸
۵–۳–۳ روش بوت‌استرپ جامعه مصنوعی گروس ۱۲۲
۵–۴ بوت‌استرپ نمونه معکوس در حالت M نامعلوم: برآوردگر مورتی ۱۲۴
۵–۵ فاصله اطمینان درستنمایی تجربی در نمونه‌گیری معکوس ۱۲۶
۵–۵–۱ درستنمایی تجربی در حالت M معلوم: برآوردگر طبقه‌بندی پسین ۱۲۷
۵–۵–۲ درستنمایی تجربی در حالت M نامعلوم: برآوردگر مورتی ۱۳۰
۵–۶ مطالعه شبیه‌سازی (داده‌های نیروی کار) ۱۳۲

فصل ششم فاصله‌های اطمینان ناپارامتری در نمونه‌گیری خوشه‌ای سازوار ۱۳۸
۶–۱ مقدمه ۱۳۸
۶–۲ بوت‌استرپ در نمونه‌گیری خوشه‌ای سازوار ۱۳۹
۶–۲–۱ روش پیشنهادی ۱ (جامعه مصنوعی) ۱۴۳
۶–۲–۲ تعیین اندازه زیرنمونه در روش جامعه مصنوعی ۱۴۴
۶–۲–۳ روش پیشنهادی ۲ (بوت‌استرپ باجایگذاری، BWR) ۱۵۱
۶–۲–۴ تعیین اندازه زیرنمونه در روش BWR ۱۵۳

۶–۳ فاصله اطمینان درستنمایی تجربی در نمونه‌گیری خوشه‌ای سازوار ۱۵۶
۶–۴ مطالعه شبیه‌سازی ۱۶۱

پیوست الف (اثبات قضیه همگرایی تابع لگاریتم درستنمایی تجربی) ۱۷۲

پیوست ب (برنامه‌های نرم‌افزاری) ۱۷۶

واژه‌نامه فارسی به انگلیسی ۱۸۲

واژه‌نامه انگلیسی به فارسی ۱۸۵

مراجع ۱۸۷

چکیده

در بسیاری از بررسی‌های نمونه‌ای با جامعه‌های کمیاب سر و کار داریم. جامعه‌هایی از این قبیل به‌طور عمده در پژوهش‌های زیست‌شناسی، علوم حیاتی و جغرافیایی یافت می‌شوند. در این وضعیت‌ها به‌طور معمول استفاده از طرح‌های نمونه‌گیری سازوار منجر به دستیابی به برآوردهای کارآتری نسبت به طرح‌های نمونه‌گیری سنتی می‌شود. از طرفی برآوردگرهای ساخته‌شده بر اساس طرح‌های سازوار به‌طور عمده دارای توزیع‌های نامتقارنی بوده و لذا فاصله‌های اطمینان بر اساس تقریب نرمال از اعتبار چندانی برخوردار نخواهند بود. از این‌رو یافتن روش‌هایی که منجر به فاصله‌های اطمینان با خواص پوششی مناسبی بر اساس چنین برآوردگرهایی شود، موضوعی در خور توجه طی سال‌های اخیر بوده است.

در این رساله با در نظر گرفتن عمده‌ترین طرح‌های نمونه‌گیری سازوار شامل نمونه‌گیری معکوس و نمونه‌گیری خوشه‌ای سازوار به مطالعه برآوردهای فاصله‌ای تحت آن‌ها خواهیم پرداخت. تمرکز ما بر روی روش‌های ناپارامتری بوت‌استرپ و درستنمایی تجربی بوده و با ارائه و تعمیم این روش‌ها به طرح‌های سازوار یادشده کارآیی آن‌ها را با تقریب نرمال برآوردگرهای متناظر مورد سنجش قرار می‌دهیم.

رده‌بندی موضوعی: ۶۲D۰۵

کلمات کلیدی: نمونه‌گیری خوشه‌ای سازوار، نمونه‌گیری معکوس، بوت‌استرپ، درستنمایی تجربی، جامعه کمیاب، فاصله اطمینان، تقریب نرمال.

لیست جدول‌ها

۲–۱ احتمال پوشش فاصله اطمینان C_N با ضریب اطمینان ۹۰٪ برای مقادیر مختلف $BR(\hat{\theta})$ ۲۹

۳–۱ میانگین توان دوم خطای برآوردگرهای نااریب میانگین جامعه تحت طرح‌های نمونه‌گیری معکوس معمولی و تصادفی ساده. ۸۴

۳–۲ میانگین توان دوم خطای برآوردگرهای نااریب میانگین جامعه تحت طرح‌های نمونه‌گیری معکوس با قوانین توقف متغیر با $n_\circ = \frac{E(n_s)}{۳}$ و تصادفی ساده. ۸۵

۳–۳ میانگین توان دوم خطای برآوردگرهای نااریب میانگین جامعه تحت طرح‌های نمونه‌گیری معکوس با قوانین توقف متغیر با $n_\circ = \frac{۲E(n_s)}{۳}$ و تصادفی ساده. ۸۶

۴–۱ مشخصات برآوردگرهای نااریب میانگین هنسن–هرویتز، هارویتز–تامپسون و هایک تحت طرح نمونه‌گیری خوشه‌ای سازوار برای داده‌های مرغابی‌های بال آبی ۱۱۱

۵–۱ فاصله اطمینان درستنمایی تجربی ۹۰٪ برای میانگین جامعه در مثال ۳.۵ . . . . . ۱۳۲

۵–۲ مشخصات برآوردگرهای نااریب میانگین جامعه تحت طرح نمونه‌گیری معکوس معمولی برای داده‌های نیروی کار. ۱۳۳

۵–۳ نتیجه‌های فاصله‌های اطمینان ۹۰ ٪ برای میانگین جامعه داده‌های نیروی کار بر اساس روش‌های تقریب نرمال، بوت‌استرپ و درستنمایی تجربی تحت نمونه‌گیری معکوس با برآوردگر مورتی ۱۳۴

۵–۴ نتیجه‌های فاصله‌های اطمینان ۹۰ ٪ برای میانگین جامعه داده‌های نیروی کار بر اساس روش‌های تقریب نرمال، بوت‌استرپ و درستنمایی تجربی تحت نمونه‌گیری معکوس با برآوردگر طبقه‌بندی پسین ۱۳۵

۶–۱ مشخصات برآوردگرهای نااریب میانگین هنسن–هرویتز، هارویتز–تامپسون و هایک تحت طرح نمونه‌گیری خوشه‌ای سازوار برای داده‌های جامعه کمیاب ۱۶۴
۶–۲ مشخصات برآوردگرهای نااریب میانگین هنسن–هرویتز، هارویتز–تامپسون و هایک تحت طرح نمونه‌گیری خوشه‌ای سازوار برای داده‌های جامعه معمولی ۱۶۵
۶–۳ میانگین، کمینه و بیشینه اندازه زیرنمونه‌های بوت‌استرپ تحت برآوردگرهای هارویتز–تامپسون و هایک ۱۶۶
۶–۴ نتیجه‌های فاصله‌های اطمینان ۹۰ ٪ برای میانگین جامعه بر اساس روش‌های تقریب نرمال، بوت‌استرپ و درستنمایی تجربی تحت نمونه‌گیری خوشه‌ای سازوار با برآوردگر هنسن–هرویتز ۱۶۷
۶–۵ نتیجه‌های فاصله‌های اطمینان ۹۰ ٪ برای میانگین جامعه بر اساس روش‌های تقریب نرمال، بوت‌استرپ و درستنمایی تجربی تحت نمونه‌گیری خوشه‌ای سازوار با برآوردگر هارویتز–تامپسون ۱۶۸
۶–۶ نتیجه‌های فاصله‌های اطمینان ۹۰ ٪ برای میانگین جامعه بر اساس روش‌های تقریب نرمال، بوت‌استرپ و درستنمایی تجربی تحت نمونه‌گیری خوشه‌ای سازوار با برآوردگر هایک ۱۶۹

فصل ۱

مقدمه

در پیمایش‌های پیچیده اغلب طرح نمونه‌گیری منجر به داده‌هایی می‌شود که از خاصیت‌های مطلوبی نظیر استقلال و هم‌توزیعی برخوردار نیستند. ساختار پیچیده جامعه نظیر طبقه‌بندی و خوشه‌بندی جامعه به‌شکل آشیانه‌ای و یا نحوه انتخاب نمونه مانند انتخاب به‌صورت بدون جایگذاری و یا بر اساس احتمال‌های نابرابر از عمده‌ترین عوامل اثرگذار در پیچیدگی پیمایش‌های نمونه‌ای هستند. با وجود مبانی نظری پایه‌گذاری‌شده بر اساس مشاهده‌های مستقل و هم‌توزیع که منجر به استنباط‌های آماری نظیر ساختن فاصله‌های اطمینان و یا آزمون فرض‌های کلاسیک می‌شود، توسعه این اصول به طرح‌های پیچیده بسیار دشوار است. علاوه بر این در بسیاری از موارد نظریه‌های مجانبی بر اساس شرایطی از جامعه و نمونه طراحی می‌شوند که اعتبار آن‌ها به‌خصوص در مورد نمونه‌های با اندازه متوسط به پایین زیر سؤال می‌رود. از سوی دیگر به‌طور معمول در نمونه‌گیری از جامعه متناهی استفاده از استنباط مدل—پایه علاوه بر نادیده پنداشتن همبستگی بین داده‌ها، همراه با محدودیت‌هایی از قبیل وجود اطلاعات کمکی در مورد شکل توزیع داده‌ها، مدل‌های مولد آن‌ها و یا تشخیص مدل مناسب خواهد بود. چنین اطلاعاتی در بسیاری از بررسی‌های نمونه‌ای در دست نبوده و یا تهیه آن مستلزم هزینه‌های زمانی و مالی بالایی خواهد بود. از این‌رو طی سه دهه اخیر فراهم آوردن روش‌های زیرنمونه‌گیری که با حفظ خواص طرح—پایه برآوردگرها بتواند ابزاری دست‌کم به قدرت‌مندی روش‌های پارامتری فراهم آورد، مورد توجه خاص پژوهش‌گران بوده است. روش‌هایی نظیر جک‌نایف، تکرار مکرر متعادل، بوت‌استرپ و درستنمایی تجربی از عمده‌ترین این روش‌ها محسوب می‌شوند.

امروزه پژوهش‌گران آماری بر این اعتقادند که در مطالعه یک جامعه لزومی به مشاهده تمام اعضای آن نمی‌باشد، بلکه در اغلب موارد مشاهده تنها بخشی از جامعه با عنوان نمونه، پژوهش‌گر را به مقصود می‌رساند. نظریه نمونه‌گیری از جامعه متناهی بر این اصل استوار است که چنانچه از مناسبت‌های آماری صحیح استفاده شود، امکان تعمیم نتیجه‌ها و اطلاعات به دست‌آمده از نمونه به جامعه اصلی وجود خواهد داشت. لذا یکی از مسایل حائز اهمیت در نمونه‌گیری، تطابق مجموعه نمونه با جامعه کل می‌باشد. به عبارتی نمونه استخراج‌شده باید معرف و نماینده مناسبی برای نمایش کل جامعه باشد. از این ایده که در دهه سوم از قرن بیستم میلادی معرفی شد، تحت عنوان روش نمایش‌دهنده [۱] یاد می‌شود. در این نظریه نقش توزیع تصادفی به‌عنوان پیوندی بین طرح و استنباط، استفاده از برآوردگرهای وزنی متناسب با احتمال و تضمین نااریبی تقریبی مربوط به توزیع نمونه‌ای در انتخاب شیوه پیمایش و ایجاد یک معیار عملی قابل قبول به‌ویژه در متن آمارهای رسمی از جایگاه بسیار مهمی برخوردار می‌باشند.

در حالت کلی برای نمونه‌گیری دو روش نمونه‌گیری احتمالی و غیر احتمالی معرفی شده است. در نمونه‌گیری غیر احتمالی بحث تعمیم یافته‌ها به جامعه مورد بررسی مطرح نیست و چون به واحدهای منتخب احتمال‌هایی نسبت داده نمی‌شود، تعیین دقت نمونه‌گیری میسر نخواهد بود. در این روش با توجه به محدودیت جمع‌آوری اطلاعات به‌ویژه در پژوهش‌های شبه تجربی مانند بررسی‌های بالینی، جهت افزایش دقت و اعتبار، اقدام به نمونه‌گیری از جمعیت قابل دسترس می‌شود. از جمله متداول‌ترین روش‌های نمونه‌گیری غیر احتمالی می‌توان به دو روش نمونه‌گیری سهمیه‌ای و قضاوتی اشاره نمود. نمونه‌گیری احتمالی [۲] اصطلاحی است که اولین بار دمینگ در سال ۱۹۵۰ آن را به‌کار برد [۲۵]. در این راهبرد هر واحد نمونه با احتمالی مشخص و مثبت از جامعه استخراج می‌شود. امروزه تقریباً در تمام کاربردها این روش جایگزین نمونه‌گیری غیر احتمالی شده است. طرح‌هایی نظیر تصادفی ساده، طبقه‌ای، خوشه‌ای، سیستماتیک و معکوس همگی از جمله طرح‌های نمونه‌گیری احتمالی می‌باشند.

در تعیین یک طرح نمونه‌گیری مؤثر که منجر به یک استنباط مناسب بر روی پارامترهای جامعه شود، عوامل متعددی دخالت دارند. بدون تردید از مهم‌ترین عوامل در این تصمیم‌گیری می‌توان از ساختار جامعه مشتمل بر شکل توزیع مقادیر متغیر تحت مطالعه و نیز موقعیت مکانی و جغرافیایی عناصر در آن یاد نمود. یکی از شکل‌های خاص جامعه که امروزه در بسیاری از مطالعه‌های پیمایشی با آن سر و کار داریم، جامعه کمیاب یا نادر [۳] است. طبق تعریف یک جامعه را کمیاب گوییم، هرگاه از نظر دارا بودن ویژگی خاص به دو گروه با ویژگی‌های کاملاً متفاوت تقسیم شود، طوری که در گروه اول تنها تعداد کمی از

[۱] Representative Method
[۲] Probability Sampling
[۳] Rare Population

واحدهای دارای صفت مورد نظر قرار گرفته (گروه کمیاب) و بخش عمده‌ای از عناصر جامعه که فاقد این ویژگی هستند در گروه دوم جای دارند. به‌طور عمده ویژگی خاص یادشده بر حسب مقادیر متغیر تحت مطالعه تعیین می‌شود؛ مقدار متغیر تحت مطالعه در گروه کمیاب نسبتاً بزرگ و در گروه دیگر بسیار کوچک (برابر یا نزدیک صفر) است. در حالتی خاص اگر عناصر زیرگروه کمیاب جامعه در قالب تعدادی خوشه در کنار یکدیگر قرار گرفته باشند، برای آن از اصطلاح جامعه خوشه‌ای کمیاب [۴] استفاده می‌شود.

جامعه‌های کمیاب به‌طور عمده در مطالعه‌های زیست‌شناسی، علوم حیاتی و جغرافیایی یافت می‌شوند. در این وضعیت‌ها برآورد پارامترهای جامعه، نظیر میانگین یا مجموع کل واحدها بر اساس طرح‌های نمونه‌گیری سنتی با اندازه نمونه‌ای ثابت، چندان معقول به‌نظر نمی‌رسد. از این‌رو در چنین شرایطی به‌طور معمول از طرح‌های نمونه‌گیری سازوار که از کارآیی بیشتری نسبت به طرح‌های سنتی برخوردارند، استفاده می‌گردد. برخلاف طرح‌های نمونه‌گیری سنتی که در آن‌ها مشاهده مقادیر متغیر تحت مطالعه برای واحدهای نمونه پس از اتمام فرآیند نمونه‌گیری و انتخاب نمونه نهایی نیز امکان‌پذیر است، در طرح‌های نمونه‌گیری سازوار نمونه نهایی به شکل دنباله‌ای در طی فرآیند نمونه‌گیری به دست آمده و انتخاب عناصر جامعه به موضع یا مقادیر متغیر تحت مطالعه عناصر انتخاب‌شده در نمونه بستگی دارد. بدین ترتیب یکی از پیامدهای مهم در انواع طرح‌های نمونه‌گیری سازوار تصادفی بودن اندازه نمونه‌های تولیدشونده خواهد بود. به‌طور مختصر دو دلیل عمده برتری نسبی طرح‌های سازوار نسبت به طرح‌های سنتی را می‌توان در دو مورد کلی زیر بیان نمود:

۱ — در انتخاب نمونه به‌صورت سنتی این احتمال وجود دارد که در نمونه انتخاب‌شده هیچ واحدی از گروه کمیاب حضور نداشته و در نتیجه برآوردی غیر واقعی و نامطلوب از پارامترها به‌دست آید.

۲ — قرار گرفتن عناصر دارای صفت خاص در قالب تعدادی خوشه در مجاورت یکدیگر موجب می‌شود تا با دسترسی به یک عنصر از هر خوشه بتوان با کم‌ترین هزینه به همه عناصر واجد شرایط در آن خوشه دست یافته و تعداد عناصر دارای این ویژگی را در نمونه افزایش داد.

در راستای تحقق اولین مزیت برخواسته از دلیل ۱ می‌توان به طرح نمونه‌گیری معکوس اشاره نمود. در این طرح، فرآیند نمونه‌گیری تا مشاهده تعداد مشخصی از واحدهای دارای صفت کمیاب در نمونه ادامه می‌یابد. بنابراین اندازه نمونه یک متغیر تصادفی است که مقدار آن بعد از اتمام عملیات نمونه‌گیری به‌دست خواهد آمد. نمونه‌گیری معکوس که گاهی آن را نمونه‌گیری دنباله‌ای [۵] نیز می‌نامند، در حالت‌های بدون جایگذاری و با جایگذاری و نیز با احتمال انتخاب برابر و نابرابر مورد بررسی قرار گرفته است.

---

[۴] Clustered Rare Population

[۵] Sequential Sampling

دومین مزیت طرح‌های سازوار را می‌توان با اجرای یک طرح نمونه‌گیری خوشه‌ای سازوار تحقق بخشید. بر خلاف نمونه‌گیری معکوس که دارای پیشینه درازمدتی است، سابقه معرفی و به‌کارگیری نمونه‌گیری خوشه‌ای سازوار به دو دهه اخیر توسط تامپسون [۹۳] برمی‌گردد. در این طرح که توأم با تعریف همسایگی و شرط لازم جهت حضور اعضای جامعه در نمونه است، علاوه بر امکان ترکیب طرح با سایر طرح‌های نمونه‌گیری سنتی، تنوع در به‌کارگیری واحدهای نمونه در مرحله برآورد، امکان تعریف برآوردگرهای متفاوتی را فراهم می‌آورد. از جمله متداول‌ترین برآوردگرها در نمونه‌گیری خوشه‌ای سازوار می‌توان به برآوردگرهای اصلاح‌شده هنسن—هرویتز و هارویتز—تامپسون اشاره نمود. دسته دیگری از برآوردگرها تحت عنوان برآوردگر مورتی نیز وجود دارند که توسط صالحی و سیبر برای حالت‌های مختلفی از نمونه‌گیری سازوار اعم از طرح‌های نمونه‌گیری معکوس و خوشه‌ای سازوار با انتخاب نمونه اولیه به‌صورت معکوس توسعه یافته است [۷۹، ۸۰].

با توسعه قابل توجه طرح‌ها و برآوردگرهای متناظر در حیطه نمونه‌گیری سازوار انواع مسائل متنوعی که در حیطه آمار کلاسیک مطرح شده و مورد بررسی قرار گرفته‌اند، برای این دسته از طرح‌ها و برآوردگرهای ارائه‌شده تحت آن‌ها به صورت باز مورد توجه‌اند. از جمله مهم‌ترین این مسائل می‌توان به مسأله یافتن برآوردگرهای مطلوبی برای پارامترهایی به شکل غیر خطی از مقادیر متغیر تحت مطالعه اشاره نمود. از این مهم‌تر آن‌که برآوردگرهای ساخته‌شده بر اساس طرح‌های نمونه‌گیری سازوار به دلیل ماهیت کمیاب و یا خوشه‌ای جامعه به‌طور عمده از توزیع‌های نامتقارنی برخوردار می‌باشند. از این‌رو فاصله‌های اطمینان بر اساس تقریب نرمال توزیع برآوردگرهای متناظر از اعتبار چندانی برخوردار نخواهند بود. لذا یافتن روش‌هایی برای ساختن فاصله اطمینان با خواص پوششی مناسب بر اساس چنین برآوردگرهایی موضوعی در خور توجه طی سال‌های اخیر بوده است.

از جمله متداول‌ترین روش‌های ناپارامتری که به‌عنوان راهکاری مناسب برای ساختن فاصله‌های اطمینان در چند دهه اخیر مورد توجه ویژه پیمایش‌گران آماری قرار گرفته انواع روش‌های زیرنمونه‌گیری است که به‌عنوان یکی از مهم‌ترین دسته از آن‌ها می‌توان به انواع روش‌های بوت‌استرپ [۶] اشاره نمود. ویژگی بارز اینگونه روش‌ها حفظ الگو و ساختار پیچیده جامعه در زیرنمونه‌های به دست‌آمده، و در نتیجه در مرحله مهم مدل‌سازی داده‌های پیمایشی است. به‌طور کلی از روش‌های بوت‌استرپ زمانی که توزیع برآوردگر مورد علاقه نامعلوم و یا پیچیده باشد استفاده می‌شود. با این وجود، نظریه مربوط به این روش‌ها تنها در حالت‌های نمونه‌گیری با اندازه ثابت توسعه یافته و نظریه‌ای مبتنی بر طرح‌های نمونه‌گیری سازوار که منجر به اندازه‌های نمونه‌ای تصادفی می‌شوند در دست نیست.

[۶] Bootstrap

از سوی دیگر یکی از روش‌های ناپارامتری که با معرفی آن به‌طور کلاسیک در سال ۱۹۸۸ برای مشاهده‌های مستقل و هم‌توزیع توسط اوون [٦١]، به‌طور گسترده‌ای در سال‌های اخیر در حوزه نمونه‌گیری با احتمال‌های نابرابر نیز مورد توجه قرار گرفته، روش درستنمایی تجربی [۷] است. چنین روشی به ما اجازه می‌دهد تا صرف‌نظر از نوع توزیع داده‌های جمع‌آوری‌شده به برآوردهای نقطه‌ای و یا فاصله‌ای مناسبی برای پارامترهای مورد نظر جامعه بر اساس بیشینه کردن تابع توزیع تجربی داده‌ها دست یابیم. در این روش توزیع تجربی مشاهده‌ها از طریق بیشینه کردن تابع لگاریتم درستنمایی تجربی تعریف می‌شود. این راهکار توسط چن و کین [١٦] در سال ۱۹۹۳ برای نمونه‌گیری از جامعه متناهی توسعه داده شده و اصلاحاتی روی تابع درستنمایی تجربی معرفی‌شده توسط وو و رائو [۹۹] صورت پذیرفته است.

در این رساله به مطالعه روش‌های ناپارامتری بوت‌استرپ و درستنمایی تجربی در حیطه نمونه‌گیری سازوار می‌پردازیم. به‌طور خاص دو دسته عمده از طرح‌های نمونه‌گیری سازوار شامل طرح‌های نمونه‌گیری معکوس و خوشه‌ای سازوار را در نظر خواهیم گرفت. در ادامه این فصل ابتدا تعریفی از مفاهیم مورد نیاز در فصل‌های بعدی ارائه می‌شود. در فصل دوم مروری اجمالی بر روش‌های مرسوم جهت ساختن برآوردهای فاصله‌ای در حوزه نمونه‌گیری پیمایشی با تمرکز روی طرح‌های سنتی خواهیم داشت. این روش‌ها به‌طور عمده شامل تقریب نرمال به‌عنوان یک روش پارامتری و روش‌های زیرنمونه‌گیری بوت‌استرپ و درستنمایی تجربی به‌عنوان دو روش ناپارامتری خواهند بود. انواع طرح‌های نمونه‌گیری معکوس با احتمال انتخاب برابر و نیز نمونه‌گیری خوشه‌ای سازوار در فصل‌های ۳ و ۴ معرفی خواهند شد. در فصل‌های پنجم و ششم نیز روش‌های ناپارامتری بوت‌استرپ و درستنمایی تجربی را به طرح‌های یادشده تعمیم داده و دقت این روش‌ها را مورد ارزیابی قرار می‌دهیم. بیشترین تمرکز ما در این رساله بر روی پارامتر میانگین جامعه خواهد بود. این مطلب به دو دلیل عمده زیر است:

الف— اهمیت بالا و کاربرد زیاد میانگین جامعه در انواع بررسی‌های نمونه‌ای.

ب— تعریف‌پذیری اکثر پارامترهای جامعه، نظیر تغییرات کل، نسبت دو مجموع کل و ... به‌صورت توابعی خطی یا غیر خطی از میانگین جامعه.

[۷] Empirical Likelihood

۱—۱ تعریف‌ها و نمادها

در این بخش به معرفی برخی اصطلاحات و نمادهایی می‌پردازیم که در فصل‌های بعدی این رساله مورد استفاده قرار خواهند گرفت. شکل این نمادها را به‌گونه‌ای در نظر گرفته‌ایم که به‌طور معمول در بیشتر متون نمونه‌گیری مورد استفاده قرار می‌گیرند.

تعریف ۱.۱ **جامعه متناهی**[۸] : مجموعه‌ای مشتمل بر تعداد متناهی عناصر متمایز تشکیل یک جامعه متناهی می‌دهند. به‌طور معمول یک جامعه متناهی را با U و تعداد عناصر آن را با N نمایش می‌دهیم. به هر عنصر مانند u از جامعه U یک عدد به عنوان برچسب اختصاص می‌یابد. لذا:

$$U = \{u_۱, u_۲, \ldots, u_N\}$$

تعریف ۲.۱ **نمونه**[۹] : زیرمجموعه‌ای از عناصر جامعه که بر اساس روش مشخصی انتخاب شده و ویژگی‌هایی از آن‌ها اندازه‌گیری می‌شوند، تشکیل یک نمونه می‌دهد. این زیرمجموعه به‌طور معمول با s نشان داده شده و تعداد عناصر آن که اندازه نمونه نامیده می‌شود را با $n(s)$ نمایش می‌دهیم. دو تعریف برای اصطلاح نمونه وجود دارد که در اکثر مواقع مورد استفاده قرار می‌گیرد:

الف) دنباله‌ای متناهی به‌صورت $\{k_۱, k_۲, \ldots, k_{n(s)}\}$ که به ازای هر $i = ۱, ۲, \ldots, n(s)$ داریم: $k_i \in U$.

ب) زیرمجموعه‌ای غیر تهی از U شامل n عنصر.

حالت (الف) که در آن عناصر انتخاب‌شده لزوماً مجزا نیستند منطبق بر نمونه‌گیری با جایگذاری، و حالت (ب) که واحدهای انتخاب‌شده در آن مجزا هستند معادل نمونه‌گیری بدون جایگذاری می‌باشند. در صورتی که اندازه نمونه از قبل معلوم و ثابت باشد، از نماد ساده‌تر n برای نمایش اندازه نمونه استفاده می‌شود. برای هر تعریف مفروض از نمونه، S را معرف مجموعه تمام نمونه‌های ممکن (با احتمال انتخاب مثبت) از جامعه U در نظر می‌گیریم.

در این‌جا تمایز بین عنصر و واحد نمونه‌گیری بسیار حائز اهمیت است. عناصر اجزای تشکیل‌دهنده جامعه‌اند، در حالی که واحدهای نمونه‌گیری مجموعه‌هایی ناهمپوشا از عناصر جامعه‌اند که در مرحله نمونه‌گیری مورد انتخاب قرار می‌گیرند.

---

[۸] Finite Population

[۹] Sample

تعریف ۳.۱ اندازه احتمال [۱۰] : تابعی نامنفی مانند p بر روی فضای $\mathcal{S}$ است که برای هر نمونه ممکن s به‌صورت $p(s) = P_r(\mathcal{S} = s)$ تعریف می‌شود، به گونه‌ای که

$$\sum_{s \in \mathcal{S}} p(s) = ۱$$

همچنین زوج $(\mathcal{S}, p)$ یا تابع احتمال $\{p(s), s \in \mathcal{S}\}$ را یک طرح نمونه‌گیری می‌نامیم. در یک طرح نمونه‌گیری احتمال انتخاب هر زیرمجموعه از $\mathcal{S}$ به‌عنوان یک نمونه محاسبه می‌شود.

تعریف ۴.۱ نشانگر عضویت نمونه [۱۱] : فرض کنید $s \ni k$ پیشامدی را نشان دهد که در آن نمونه s، واحد k–اُم جامعه را شامل شود. در این صورت متغیر تصادفی دو وضعیتی $I_k(s)$ که با عبارت زیر تعریف می‌شود را نشانگر عضویت نمونه گویند:

$$I_k(s) = \begin{cases} ۱ & if \quad k \in s \\ ۰ & Otherwise. \end{cases}$$

قابل توجه است که نماد نشانگر فوق را صرفاً به‌منظور استفاده در حیطه نمونه‌گیری پیمایشی تعریف نمودیم. مشابه این تعریف تابع دیگری با عنوان تابع نشانگر نیز وجود دارد که در این متن از آن استفاده خواهیم نمود. فرض کنید A یک پیشامد دلخواه و $I(A)$ تابع نشانگر متناظر آن باشد. در این صورت اگر پیشامد A رخ دهد، آنگاه $I(A) = ۱$، و در غیر این‌صورت $I(A) = ۰$.

تعریف ۵.۱ احتمال شمول مرتبه اول [۱۲] : به ازای هر واحد جامعه، احتمال شمول مرتبه اول برابر با مقدار عددی احتمال انتخاب شدن آن واحد در نمونه است. این احتمال با نماد π_k نمایش داده شده و با معادله زیر تعریف می‌شود:

$$\pi_k = \sum_{s \ni k} p(s) = P_r(I_k(s) = ۱),$$

که در آن $\sum_{s \ni k}$ مجموع روی تمام نمونه‌های ممکن s شامل واحد k–اُم می‌باشد. هر طرح نمونه‌گیری که در آن احتمال‌های شمول همه عناصر برابر باشند، طرح خود–وزن نامیده می‌شود. همچنین اگر نمونه‌گیری بر اساس طرحی اجرا شود که در آن همه عناصر جامعه از احتمال شمول مثبت برخوردار باشند، از آن تحت عنوان نمونه‌گیری احتمالی یاد می‌شود. مزیت عمده این‌گونه طرح‌ها آن است که امکان دستیابی به برآوردهایی نااریب از کمیت‌های خطی برحسب مقادیر متغیر تحت مطالعه را فراهم می‌سازد.

[۱۰] Probability Measure
[۱۱] Sample Membership Indicator
[۱۲] First-order Inclusion Probability

تعریف ۶.۱ احتمال شمول مرتبه دوم [۱۳] : به ازای هر دو واحد دلخواه k, l، احتمال شمول مرتبه دوم برابر با احتمال انتخاب توأم این دو واحد در نمونه بوده و با π_{kl} نمایش داده می‌شود. به‌عبارتی

$$\pi_{kl} = \sum_{s \ni (k,\, l)} p(s) = P_r(I_k(s) = ۱، I_l(s) = ۱)$$

مشابه تعریف قبل مجموع $\sum_{s \ni (k,\, l)}$ روی تمام نمونه‌های ممکن شامل واحدهای k–اُم و l–اُم در نظر گرفته می‌شود. بدیهی است که $\pi_{kk} = \pi_k$. طرحی که در آن به ازای هر دو عنصر متفاوت k و l داشته باشیم: $\pi_{kl} > ۰$، طرح اندازه‌پذیر نامیده می‌شود. خاصیت مطلوب چنین طرح‌هایی امکان یافتن برآوردهایی نااریب از واریانس برآوردگرهای خطی است.

تعریف ۷.۱ متغیرهای پیمایشی [۱۴] : مجموعه‌ای از متغیرهای مورد علاقه که مقادیر آن‌ها در یک پیمایش اندازه‌گیری شده و یا در مراحل مختلف نمونه‌گیری مورد استفاده قرار می‌گیرد، متغیرهای پیمایشی یا به عبارتی متغیرهای تحت مطالعه نامیده می‌شوند. در این رساله تنها متغیرهای یک‌بُعدی را در نظر گرفته و از حرف y برای نمایش آن استفاده می‌کنیم. همچنین مقدار این متغیر برای عنصر k–اُم جامعه با y_k نشان داده می‌شود.

تعریف ۸.۱ پارامتر جامعه متناهی [۱۵] : به هر تابع حقیقی از مقادیر $y_۱, y_۲, \ldots, y_N$ یک پارامتر گویند. پارامترهایی که به طور عمده در این رساله با آن‌ها سر و کار خواهیم داشت، عبارتند از:

میانگین جامعه: $\bar{y}_U = \frac{۱}{N}\sum_{k=۱}^{N} y_k$ واریانس کل جامعه: $\sigma_y^۲ = \frac{N-۱}{N} S_y^۲ = \frac{۱}{N}\sum_{k=۱}^{N}(y_k - \bar{y}_U)^۲$

به‌طور معمول بیشتر پارامترها را می‌توان بر حسب میانگین کل متغیرهای موجود تعریف نمود. در این میان برخی از پارامترها به شکل توابعی خطی و برخی به صورت غیر خطی از میانگین‌های کل ظاهر می‌شوند. به‌عنوان مثالی از کمیت‌های خطی می‌توان به تابع توزیع تجمعی مشاهدات جامعه با ضابطه

$$F(y) = \frac{۱}{N}\sum_{k=۱}^{N} I(y_k \le y)$$

اشاره نمود. نمونه‌های متعددی از کمیت‌های غیر خطی نیز وجود دارد که از معروف‌ترین آن‌ها می‌توان از چندک p–اُم جامعه با تعریف $Q(p) = F^{-۱}(p)$ و یا نسبت دو میانگین کل با ضابطه $R_{yx} = \frac{\bar{y}_U}{\bar{x}_U}$ نام برد.

---

[۱۳] Second-order Inclusion Probability

[۱۴] Survey Variables

[۱۵] Finite Population Parameter