دانشکده علوم ریاضی

رساله

برای دریافت درجه دکتری در رشته
ریاضی ، گرایش آمار-نمونه‌گیری

عنوان

متغیرهای کمکی در نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار

اساتید راهنما

دکتر محمد صالحی و دکتر افشین پرورده

استاد مشاور

دکتر علیرضا نعمت اللهی

پژوهشگر

بردیا پناه‌بحق

۱۳۹۱

نام خانوادگی دانشجو: پناه‌بحق نام: بردیا

عنوان: متغیرهای کمکی در نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار

اساتید راهنما: دکتر محمد صالحی و دکتر افشین پرورده
استاد مشاور: دکتر علیرضا نعمت اللهی

مقطع تحصیلی: دکتری رشته: ریاضی گرایش: آمار- نمونه‌گیری

دانشگاه: دانشگاه صنعتی اصفهان دانشکده علوم ریاضی
تاریخ فارغ‌التحصیلی: ۱۳۹۱ تعداد صفحات: ۱۲۱

واژگان کلیدی: نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار، براوردگرهای رگرسیونی و نسبتی، بهینه‌سازی، نمونه‌گیری دوگانه، خواص حدی در جامعه‌های متناهی، جامعه‌های کمیاب

چکیده

در این رساله سعی شده‌است که نقش متغیرهای کمکی در هردو مرحله جمع‌آوری نمونه و براورد پارامتر مجهول جامعه در چند طرح سازوار، با تمرکز بر نمونه‌گیری دنباله‌ای دومرحله‌ای، مورد بررسی قرار گیرد. برای این منظور ابتدا فرض معلوم بودن میانگین متغیر کمکی در کل جامعه را بر روی مسئله می‌گذاریم و پس از آن، با حذف این فرض و با تلفیق نمونه‌گیری دوگانه و نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار، طرحی برای استفاده از متغیرهای کمکی در جامعه‌های کمیاب و خوشه‌ای، بدون نیاز به اطلاع پیشین از میانگین جمعیتی متغیرهای کمکی، ارائه می‌دهیم. در این حالت برای منصفانه بودن مقایسه طرح‌ها، یک تابع هزینه برای طرح‌ها تعریف و محاسبه می‌کنیم. در طی این مباحث، قدرت کشف طرح‌های سازوار در مقایسه با طرح‌های سنتی نیز محاسبه و بحث می‌شود.

در ادامه برای افزایش قدرت کشف طرح نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار، طرح را برای حالت چندشرطی در مقابل یک شرطی بسط می‌دهیم. همچنین براوردگری کارا برای این طرح پیشنهاد داده و خواص حدی این براوردگر را در طرح مربوطه مورد بررسی قرار می‌دهیم. همچنین در یکی از طرح‌های کارای سازوار با نام طرح نمونه‌گیری دنباله‌ای طبقه‌ای با تخصیص کامل، نقش متغیرهای کمکی را بسط داده و بررسی می‌کنیم. در هریک از قسمت‌های قبل، نتایج با استفاده از شبیه‌سازی‌ها و چند مجموعه از داده‌های واقعی مورد بررسی و ارزیابی دقیق‌تر قرار می‌گیرند. در پایان نیز قسمتی از برنامه‌های نوشته شده توسط نرم‌افزار R برای استفاده پژوهشگران ارائه می‌شود.

كلية حقوق اعم از چاپ و تكثير، نسخه برداري ، ترجمه، اقتباس و ... از اين پايان نامه براي دانشگاه صنعتی اصفهان محفوظ است. نقل مطالب با ذكر مأخذ آزاد است.

فهرست مطالب

۱ مقدمه ۲

۱.۱ تاریخچه موضوعی ۳

۲.۱ فصل‌بندی رساله ۵

۲ متغیرهای کمکی در نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار ۷

۱.۲ نقش متغیرهای کمکی در طراحی نمونه‌گیری در ATS ۷

۱.۱.۲ طرح و براوردگر مربوطه ۸

۲.۱.۲ شبیه‌سازی‌ها ۱۰

۳.۱.۲ نتایج ۱۱

۴.۱.۲ جمع بندی و پیشنهادات ۱۲

۲.۲ متغیرهای کمکی در مرحله‌ی براورد در نمونه‌گیری دنباله‌ای دومرحله‌ای به صورت براوردگر رگرسیونی ۱۵

۱.۲.۲ طرح و براوردگر رگرسیونی متناسب با طرح ۱۷

۳.۲ متغیرهای کمکی در مرحله‌ی براورد در نمونه‌گیری دنباله‌ای دومرحله‌ای به صورت براوردگر نسبتی ۲۲

۱.۳.۲ براوردگر نسبتی مناسب برای طرح ATS ۲۵

۴.۲ امید مشاهده تعداد نقاط کمیاب تحت برخی طرح‌های سازوار و مرسوم ۲۷

۵.۲ شبیه‌سازی‌ها ۲۹

۱.۵.۲ صدف‌های کمیاب ۳۰

۲.۵.۲ جامعه پوآسن خوشه‌ای ۳۲

۳.۵.۲ مطالعه فراوانی گیاهان بر اساس اطلاعات نقشه‌های فضایی ۳۳
۴.۵.۲ چند جامعه جدید با رابطه ناخطی بین دو متغیر ۳۶
۶.۲ جمع بندی و پیشنهادات ۳۷

۳ نمونه‌گیری دنباله‌ای دومرحله‌ای دوگانه با تابع هزینه ۴۷
۱.۳ نمادها و طرح نمونه‌گیری ۴۷
۲.۳ براوردگر رگرسیونی مناسب ۴۸
۳.۳ امید و واریانس تقریبی براوردگر رگرسیونی ۴۹
۴.۳ محاسبه‌ی $\hat{\beta}_o$ ۵۳
۵.۳ براوردگر نسبتی مناسب ۵۵
۶.۳ شبیه‌سازی‌ها ۵۷
۱.۶.۳ تعریف تابع هزینه طرح و امید آن ۵۸
۷.۳ جمع بندی و پیشنهادات ۶۲

۴ بررسی خواص حدی براوردگری کارا در نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار چند شرطی ۶۸
۱.۴ طرح و براوردگر مربوطه ۶۹
۲.۴ باز نویسی براوردگر مورتی و تبدیل آن به صورتی جدید ۷۱
۳.۴ خواص حدی میانگین نمونه در طرح نمونه‌گیری دنباله‌ای طبقه‌ای سازوار با چند شرط ($MASS$) ۷۱
۱.۳.۴ براوردگر در طرح کامل $MASS$ ۸۱
۴.۴ شبیه‌سازی‌ها ۸۳
۱.۴.۴ طرح $MATS$ ۸۳
۲.۴.۴ بررسی نرمال بودن براوردگر جدید در طرح $MASS$ ۸۹
۵.۴ جمع بندی و پیشنهادات ۹۰

۵ متغیرهای کمکی در نمونه‌گیری دنباله‌ای طبقه‌ای با تخصیص کامل ۹۳
۱.۵ طرح و براوردگرهای مربوطه ۹۴
۲.۵ شبیه‌سازی‌ها ۹۸

۱.۲.۵ فقط یک متغیر با جامعه گل‌های لاله ۹۸
۲.۲.۵ حضور متغیر کمکی ۱۰۰
۳.۵ جمع‌بندی و پیشنهادات ۱۰۲

۶ پیوست (برنامه نوشته‌شده توسط نرم افزار R برای تولید جامعه‌ای خوشه‌ای و نایاب) **۱۰۵**

واژه‌نامه فارسی به انگلیسی **۱۱۴**

واژه‌نامه انگلیسی به فارسی **۱۱۷**

لیست تصاویر

۱.۲ چهار جامعه شبیه‌سازی شده از صدف‌ها ۱۳

۲.۲ نمودار نتایج مربوط به شبیه‌سازی جامعه‌ی صدف‌ها ۱۴

۳.۲ جامعه پواسون خوشه‌ای ۳۳

۱.۴ سه جامعه‌ی شبیه‌سازی شده از صدف‌ها ۸۵

۲.۴ جامعه پواسون خوشه‌ای فقط با متغیر اصلی ۹۰

۳.۴ هیستوگرام براوردگر $\bar{y}_s$ برای جامعه۱ . . . . . ۹۱

۴.۴ هیستوگرام براوردگر $\bar{y}_s$ برای جامعه پواسون خوشه‌ای ۹۱

۵.۴ هیستوگرام براوردگر $\bar{y}_s$ برای جامعه۳ . . . . . ۹۱

۱.۵ جامعه گل‌ها لاله ۹۹

۲.۵ جامعه‌ای از صدف‌های نایاب به همراه صدف‌های کمکی ۱۰۰

لیست جداول

۱.۲ هشت جامعه شبیه‌سازی شده از صدف‌ها ۱۲

۲.۲ نتایج مربوط به اهمیت پارامترها ۱۵

۳.۲ نتایج مربوط به امید ریاضی تعداد نقاط کمیاب کشف شده برای جامعه‌ی صدف‌ها ۳۸

۴.۲ کارایی براوردگرها برای جامعه‌ی صدف‌ها برای $M = 20$ ۳۹

۵.۲ کارایی براوردگرها برای جامعه‌ی صدف‌ها برای $M = 10$ ۴۰

۶.۲ کارایی براوردگرها در حالت‌های متفاوت برای جامعه‌ی صدف‌ها برای $M = 4, 8$. ۴۱

۷.۲ نتایج مربوط به جامعه پوآسن خوشه‌ای ۴۲

۸.۲ کارایی براوردگرها برای حالت‌های مختلف در جامعه درختچه‌های گز ۴۳

۹.۲ مقایسه میزان استفاده روش‌های مرسوم و روش سازوار از اطلاعات متغیر کمکی در مرحله‌ی براورد در افزایش کارایی ۴۴

۱۰.۲ کارایی‌ها در جامعه اول ($popu1$) با رابطه‌ی ناخطی $y = 3 + sin(x) + \epsilon$ ۴۵

۱۱.۲ کارایی‌ها در جامعه دوم ($popu2$) با رابطه‌ی ناخطی $y = 10 + x^4 + \epsilon$ ۴۵

۱۲.۲ کارایی‌ها در جامعه سوم ($popu3$) با رابطه‌ی ناخطی $y = \frac{1}{x} + log(x) + \epsilon$ ۴۶

۱.۳ جامعه‌های شبیه‌سازی شده، با y به عنوان متغیر اصلی و x و z به عنوان متغیرهای کمکی. ۵۷

۲.۳ کارایی‌ها در جامعه‌ی صدف‌ها با $d = 2, D = 15$ ۶۲

۳.۳ کارایی‌ها در جامعه‌ی جامعه‌ی صدف‌ها با $d = 2, D = 20$ ۶۳

۴.۳ کارایی‌ها در جامعه۱ با $D = 30$ ۶۴

۵.۳ کارایی‌ها در جامعه۲ (x) با $D = 30$ ۶۵

۶.۳ کارایی‌ها در جامعه۲ (z) با $D = 30$ ۶۶

۷.۳ مقایسه ضریب مرسوم و ضریب بهینه ۶۷

۱.۴ خلاصه وضعیت شرط‌ها و واحدهای موجود در نمونه و جامعه، ۶۹

۲.۴ مشخصات طرح برای هنگامی که فقط یک شرط داریم ۷۵

۳.۴ PP محاسبه شده برای دو طرح در سه جامعه شبیه‌سازی شده ۸۷

۴.۴ کارایی ATS_{2c} نسبت به ATS_{1c} برای $\bar{y}_s$ برای سه جامعه‌ی صدف‌ها ۸۸

۵.۴ کارایی $\bar{y}_s$ نسبت به براوردگر مورتی در طرح دوشرطی برای سه جامعه‌ی صدف‌ها . ۸۹

۶.۴ PP برای داده‌های فرایند خوشه‌ای پواسون ۹۰

۷.۴ کارایی ATS_{2c} نسبت به ATS_{1c} برای $\bar{y}_s$ برای داده‌های فرایند خوشه‌ای پواسون . . ۹۲

۸.۴ کارایی براوردگر جدید ($\bar{y}_s$) نسبت به براوردگر مورتی در طرح دوشرطی برای داده‌های فرایند خوشه‌ای پواسون ۹۲

۱.۵ نتایج کارایی، مربوط به گل‌های لاله ۱۰۳

۲.۵ نتایج کارایی، مربوط به جامعه‌ی صدف‌ها ۱۰۴

۳.۵ نتایج مربوط به کارایی برای جامعه ۱، برای سنجش تاثیر همبستگی بالا ۱۰۴

فصل ۱

مقدمه

طرح‌های نمونه‌گیری سازوار با نشان دادن توانایی خود در بررسی جامعه‌های کمیاب همواره مورد توجه پژوهشگران بوده‌اند. یک طرح نمونه‌گیری سازوار است اگر فرایند انتخاب نمونه به مقادیر متغیر مورد مطالعه مشاهده شده در نمونه بستگی داشته باشد. از قدیمی‌ترین این طرح‌ها که طرحی کارا برای براورد نسبت یک زیر جامعه کمیاب می‌باشد، نمونه‌گیری معکوس است که در آن تا مشاهده تعداد مشخصی از عنصرها کمیاب مورد نظر، نمونه‌گیری ادامه پیدا می‌کند. در جامعه کمیاب، معمولا مجموعه عنصرها مورد مطالعه دارای فراوانی بسیار کم هستند به طوری که طرح‌های نمونه‌گیری مرسوم قادر به بررسی دقیق آن‌ها نمی‌باشند. هنگامی که عبارت "خوشه‌ای" به جامعه‌ای کمیاب اضافه می‌شود اشاره به پراکندگی خوشه‌ای‌شکل آن‌ها در محیط‌های مورد بررسی دارد. طرح‌های سازوار بدلیل کارایی خوب خود بخصوص در مسائل زیست محیطی، طی دو دهه اخیر مورد توجه بسیاری از پژوهشگران بوده‌اند. از جدیدترین طرح‌های سازوار برای بررسی جامعه‌های کمیاب و خوشه‌ای، طرح نمونه‌گیری خوشه‌ای سازوار است که در آن بطور کلی بر اساس یک طرح نمونه‌گیری مرسوم، نمونه‌ای آغازین گرفته می‌شود و هرجا یک واحد نمونه‌گیری در شرطی معین صدق کند، تمام همسایگی‌های این واحد، تا هنگامی که دیگر هیچ همسایگی‌ای در شرط صدق نکند (معمولاً در برگیرنده اعضای کمیاب مورد نظر نباشد) نمونه‌گیری می‌شوند. دو مساله اصلی در این طرح، تعریف شرط "C" که مشخص کننده اعضای مورد نظر برای ادامه فرایند نمونه‌گیری و تعریف همسایگی برای هر واحد، می‌باشند. مساله همسایگی‌ها یکی از مسائل مهم و بحرانی در بحث نمونه‌گیری سازوار است، زیرا تعریف یک همسایگی مناسب معمولاً منجر به افزایش بی‌رویه اندازه نمونه و همچنین ورود واحدهای لبه (واحدهایی که در شرط C صدق نمی‌کنند اما در همسایگی اعضایی هستند که شرط را دارا می‌باشند و بنابراین وارد نمونه نهایی می‌شوند) می‌شود.
نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار نوع دیگری از نمونه‌گیری سازوار است که علاوه بر توانایی در جهت کشف خوشه‌های کمیاب در جامعه، از تعریف و استفاده از همسایگی و لبه‌ها بی‌نیاز است. طرح‌های دنباله‌ای به طرح‌هایی گفته می‌شود که در آن‌ها فرایند گرفتن نمونه به صورت دنباله‌ای انجام

می‌شود. طرح بدین صورت است که ابتدا جامعه به تعدادی واحد نمونه‌گیری اولیه تقسیم می‌شود و هر واحد نمونه‌گیری اولیه خود به واحدهای کوچکتری به نام واحدهای نمونه‌گیری ثانویه تقسیم می‌شود. سپس یک نمونه تصادفی از واحدهای نمونه‌گیری اولیه گرفته می‌شود و داخل هریک از این واحدها یک نمونه تصادفی آغازین گرفته می‌شود و با توجه به فراوانی عنصرهای کمیاب موجود در نمونه اولیه در یک واحد خاص، که در شرط صدق می‌کنند، یک نمونه اضافی به واحد مورد نظر تخصیص داده می‌شود. وجهه دنباله‌ای بودن طرح در این‌جا بدین معنی است که طی دنباله‌ای گرفته شدن نمونه، تعداد واحدهای نمونه‌گیری شده ثانویه در هر واحد نمونه‌گیری اولیه بسته به مشاهدات نمونه آغازین قابل تغییر است. این روش در جامعه‌های کمیاب و خوشه‌ای کارایی‌های خوبی از خود نشان داده‌است. اجرا کردن نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار در تمامی طبقات یک جامعه طبقه‌بندی شده، در اکثر مواقع، با تخصیص نمونه بیشتر به طبقات با واریانس بیشتر موجب افزایش کارایی نمونه‌گیری (طبق اصل تخصیص نیمن) به میزان قابل توجهی می‌شود.

در این رساله با توجه به این‌که در اکثر پژوهش‌ها، در کنار متغیر اصلی متغیرهای کمکی فراوانی وجود داشته و جمع‌آوری می‌شوند، سعی می‌شود که نقش این متغیرهای کمکی در مرحله‌ی طراحی نمون‌گیری (جمع‌آوری نمونه) و براورد در چند طرح مختلف مورد بررسی قرار گیرد. همچنین با توجه به این‌که در این گونه طرح‌ها انتظار می‌رود با تخصیص نمونه بیشتر به واحدهای اولیه با تعداد عنصرها کمیاب بیشتر، بتوان موجب افزایش کارایی نمونه‌گیری شد، تلاش می‌شود با استفاده از متغیرهای کمکی دارای همبستگی مناسب با متغیر کمیاب اصلی، روش‌هایی برای افزایش قدرت کشف عنصرها کمیاب و در نتیجه افزایش کارایی طرح نمونه‌گیری ارائه شود. همچنین سعی می‌شود براوردگری با کارایی و رفتار حدی مناسب در این‌گونه طرح‌ها پیدا کرده و مورد بررسی بیشتر قرار دهیم و در آخر متغیرهای کمکی در یکی از کاراترین طرح سازوار دنباله‌ای دومرحله‌ای سازوار مطرح و بررسی می‌شود.

۱.۱ تاریخچه موضوعی

در ادامه به‌طور مختصر تاریخچه‌ای از کارهای صورت گرفته در زمینه‌های مطرح شده در رساله بیان می‌شوند.

نمونه‌گیری معکوس را می‌توان به عنوان اولین نوع نمونه‌گیری سازوار در نظر گرفت. هالدان [۱۹۴۵] نمونه‌گیری معکوس را برای براورد نسبت جامعه‌های کمیاب ارائه کرد. تامپسون [۱۹۹۰] نمونه‌گیری خوشه‌ای سازوار را به عنوان روشی برای بررسی جامعه‌های کمیاب و خوشه‌ای ارائه کرد. تامپسون و سیبر [۱۹۹۶]، کریستمن [۲۰۰۰] و اسمیت و همکاران [۲۰۰۴] نشان دادند این طرح یکی از گزینه‌های مناسب برای بررسی جامعه‌های کمیاب و خوشه‌ای می‌باشد. صالحی [۱۹۹۹] تحت شرایطی نشان داد که شامل کردن لبه‌ها در نمونه، موجب افزایش کارایی نمی‌شود و اگر هم بشود مقدار آن بسیار ناچیز

است. از آن‌جا که در بعضی مواقع تعریف همسایگی کاری مشکل یا ناممکن است (مانند نمونه‌گیری از نوجوانان برای براورد درصد مصرف کنندگان مواد مخدر)، صالحی و اسمیت [۲۰۰۵] طرح نمونه‌گیری دنباله‌ای دومرحله‌ای را برای رفع این مشکل ارائه کردند که دیگر احتیاجی به تعریف همسایگی ندارد ولی قادر به جستجوی اعضای کمیاب با ساختار خوشه‌ای در جامعه می‌باشد. آن‌ها با بررسی روش پیشنهادی خود برروی سه جامعه متفاوت از لحاظ پراکندگی و ساختار خوشه‌ای، نشان دادند که روش مطرح شده کارایی خوبی در بررسی جوامع مختلف از خود نشان می‌دهد. در طرح صالحی و اسمیت بر اساس وجود یا عدم وجود عضو کمیاب، تصمیم گرفته می‌شود که نمونه‌ی اضافی به واحد اولیه تخصیص داده شود یا نه و به فراوانی آن اهمیتی داده نمی‌شود. بروان و همکاران [۲۰۰۸] با تصمیم‌گیری براساس فراوانی مورد نظر، با ارائه طرحی از نمونه‌گیری دنباله‌ای دومرحله‌ای با نام نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار با توانایی بیشتر در جهت هدایت مسیر نمونه‌گیری به طرف عنصرها کمیاب، موجب افزایش کارایی طرح مورد نظر شدند. آن‌ها نشان دادند که روش جدید از روش اولیه صالحی و اسمیت کارایی بیشتری دارد. صالحی و همکاران [۲۰۱۰] برای کنترل اندازه نهایی نمونه در طرح مورد نظر روشی ارائه کردند. آن‌ها همچنین نشان دادند که اجرای طرح نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار در تمامی طبقات یک جامعه طبقه‌بندی شده، در مواقعی که اطلاعاتی در مورد واریانس داخل طبقات در اختیار نداریم، می‌تواند بدلیل هدایت ما به‌سمت تخصیص نمونه‌ی بیشتر به طبقاتی با واریانس بیشتر به گونه‌ای تداعی کننده طرح تخصیص نیمن باشد. آن‌ها طرح را در یک مطالعه خاک شناسی در سطح کره ماه و مطالعه میزان عسل تولیدی در منطقه کردستان ایران مورد بررسی قرار دادند. مرادی و صالحی [۲۰۱۰] طرح نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار را برای پیدا کردن برآودرگرهایی بهتر مورد بررسی بیشتر قرار دادند. همچنین صالحی و بروان [۲۰۱۰] طرح نمونه‌گیری طبقه‌ای با تخصیص کامل را به عنوان یکی از کاراترین و ساده‌ترین روش‌های نمونه‌گیری سازورا ارائه ارائه دادند. در ارتباط با نقش متغیرهای کمکی در طرح‌های سازوار، فلکس-مدینا و تامپسون [۲۰۰۴] در نمونه‌گیری خوشه‌ای سازوار، طرحی مشابه نمونه‌گیری دوگانه مورد بررسی قرار دادند. دریور و چااو [۲۰۰۷] یک براوردگر نسبتی را تحت احتمالات نابرابر، برای طرح نمونه‌گیری سازوار خوشه‌ای مطرح و بررسی کردند.

چااو و همکاران [۲۰۰۸] با استفاده از قضیه رائو-بلکول براوردگر نسبتی مربوطه را از نظر واریانس بهبود بخشیدند. کاردیلار و سینجی [۲۰۰۵] براوردگر نسبتی را در یک نمونه‌گیری طبقه‌ای سازوار خوشه‌ای مورد بررسی قرار دادند. چوتیمن [۲۰۱۰] با اضافه کردن یک ضریب ثابت به براوردگر نسبتی و بهینه‌سازی آن موجب افزایش کارایی براوردگر مربوطه در طرح نمونه‌گیری طبقه‌ای سازوار خوشه‌ای شد و مرادی و همکاران [۲۰۱۱] برای بررسی میزان آرسنیک موجود در خاک، در یک مطالعه موردی، استفاده از متغیرهای کمکی و براوردگر رگرسیونی را در طرح نمونه‌گیری معکوس معرفی و بررسی کردند. خواص حدی براوردگرها در جامعه‌های متناهی نیز طی ۶۰ سال گذشته در مرکز توجه پژوهشگرانی در زمینه‌های نمونه‌گیری و احتمال بوده است. اردوش و رنی [۱۹۵۹] و هایک [۱۹۶۰] بطور مستقل نرمال بودن میانگین نمونه در یک نمونه‌گیری تصادفی ساده بدون جایگذاری را در یک جامعه متناهی

مورد بررسی قرار دادند. روزن [۱۹۷۲] براوردگر هورویتز-تامپسون را در یک نمونه‌گیری دنباله‌ای با احتمال‌های متفاوت، اسکات و وو [۱۹۸۱] براوردگرهای رگرسیونی و نسبتی را در نمونه‌گیری تصادفی ساده بدون جایگذاری و اوهلسون [۱۹۸۹] خواص حدی در طرح‌های دومرحله‌ای را مورد بررسی قرار دادند. سارندال [۱۹۸۰] و ایساکی و فولر [۱۹۸۲] سازگاری براوردگر رگرسیونی را تحت طرح‌های مختلف مورد بررسی قرار دادند. چن و رائو [۲۰۰۷] نیز نرمال بودن میانگین نمونه را در یک نمونه‌گیری دوفازی تحت شرایطی اثبات کردند. برگر [۱۹۹۸] نرخ همگرایی به توزیع نرمال را برای براوردگرهای هورویتز-تامپسون مورد بررسی قرار داد و فلکس-مدینا [۲۰۰۳] نرمال بودن براوردگر هورویتز-تامپسون را در نمونه‌گیری خوشه‌ای سازوار بررسی کرد. در نهایت فولر [۲۰۰۹] بسیاری از مطالب مربوط به قضایای حدی در جامعه‌های متناهی را در قالب یک کتاب گرداوری کرده و مورد بررسی جامعی قرار داد.

۲.۱ فصل‌بندی رساله

در این رساله سعی شده نقش متغیرهای کمکی در هردو مرحله‌ی طراحی و براورد پارامتر مجهول جامعه در چند طرح سازوار، با تمرکز بر نمونه‌گیری دنباله‌ای دومرحله‌ای، در قالب فصل‌بندی زیر مورد بررسی قرار گیرد. لازم به ذکر است که نتایج بدست آمده در فصل‌ها، اصیل می‌باشند و مقاله‌های ارجاع داده شده در ادامه فصل‌بندی، از رساله استخراج شده‌اند و در فصل‌های آینده به آن‌ها ارجاع نخواهیم داد. در فصل ۲ ابتدا نقش متغیرهای کمکی در مرحله‌ی طراحی توسط یک مجموعه از داده‌های شبیه‌سازی شده براساس داده‌های واقعی از صدف‌های در معرض انقراض آب‌های شیرین به تفصیل مورد بررسی قرار می‌گیرد. در ادامه فصل، نقش متغیرهای کمکی در مرحله‌ی براورد نیز با معرفی چند براوردگر رگرسیونی و نسبتی متناسب با طرح، با فرض معلوم بودن میانگین کل جمعیت متغیر کمکی، مطرح شده و مورد بررسی قرار می‌گیرند. در پایان فصل در قالب چند سری داده واقعی و شبیه‌سازی شده نتایج فصل مورد ارزیابی قرار خواهند گرفت (پناه‌بحق و همکاران [۲۰۱۱] و صالحی و همکاران [۲۰۱۲]).

در فصل ۳ با توجه به معلوم نبودن میانگین جمعیتی متغیرهای کمکی در بسیاری از پژوهش‌ها، طرح جدیدی که حاصل تلفیق نمونه‌گیری دوگانه با نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار است ارائه می‌شود. طرح جدید مطرح شده بدون نیاز به اطلاع قبلی از میانگین جمعیتی متغیرهای کمکی، توانایی بررسی جامعه‌های خوشه‌ای و کمیاب را دارد. با این تلفیق با یک طرح چند مرحله‌ای مواجه می‌شویم که اندکی کار محاسبه براوردگرها و واریانس آن‌ها پیچیده می‌شود. در این فصل براوردگری نااریب برای میانگین جامعه معرفی شده و واریانس تقریبی براوردگر و براوردگری برای آن محاسبه می‌شوند. در پایان این فصل در قالب چند جامعه شبیه‌سازی شده و چند جامعه واقعی، کارایی طرح در عمل مورد بررسی قرار

می‌گیرد. همچنین برای منصفانه بودن مقایسه‌ها، تابعی برای محاسبه‌ی هزینه‌ی طرح‌ها معرفی کرده و در شبیه‌سازی‌ها مورد استفاده قرار می‌دهیم.

در فصل ۴ با توجه به ایده‌ای که پشت طرح‌های دومرحله‌ای سازوار وجود دارد، یعنی تخصیص نمونه بیشتر به طبقاتی با تعداد اعضای کمیاب بیشتر (که انتظار داریم واریانس بیشتری داشته باشند)، تغییراتی در طرح (مرحله‌ی جمع‌آوری نمونه) مدنظر اعمال می‌کنیم. در این فصل بجای تعریف یک شرط برای تخصیص نمونه اضافی، چند شرط برروی متغیرهای جامعه تعریف کرده و نشان می‌دهیم در جامعه‌های مستعد، موجب افزایش قدرت کشف اعضای کمیاب جامعه می‌شود. باید توجه داشت که قدرت کشف بیشتر، به خودی خود یک شاخص مهم در جامعه‌های کمیاب می‌باشد. پس از آن با توجه به شکل براوردگر مرسوم در این طرح، یک براوردگر جدید معرفی می‌کنیم و خواص حدی این براوردگر جدید در این طرح سازوار را مورد بررسی قرار می‌هیم. بررسی خواص حدی در جامعه‌های متناهی معمولا پیچیده و طولانی می‌شوند که بررسی ما نیز از این قاعده مستثنی نبوده. بنا به رسم فصل‌های قبل در این فصل نیز توسط چند سری داده‌های شبیه‌سازی شده نتایج مورد بررسی بیشتر قرار می‌گیرند (پرورده و همکاران [۲۰۱۲]).

فصل ۵ در برگیرنده بررسی با جزئیات بیشتر یکی از جدیدترین و کاراترین طرح‌های سازوار با عنوان طرح نمونه‌گیری دنباله‌ای طبقه‌ای با تخصیص کامل می‌باشد. نقش متغیرهای کمکی در این طرح در مرحله‌ی جمع‌آوری و براورد مورد بررسی قرار می‌گیرد (بروان و همکاران[1] [۲۰۱۱]، بروان و همکاران[2] [۲۰۱۱] و بروان و همکاران [۲۰۱۲]).

در پیوست (فصل ۶) نیز، برای شبیه‌سازی جامعه‌های با ساختار خوشه‌ای یک برنامه نوشته شده توسط نرم‌افزار "R" آورده شده است.

فصل ۲

متغیرهای کمکی در نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار

با توجه به ساختار طرح نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار(ATS) ، متغیرهای کمکی می‌توانند در هر دومرحله‌ی طراحی نمونه‌گیری و مرحله‌ی براورد پارامتر مجهول جامعه مورد استفاده قرار گیرند. در این فصل تلاش می‌شود نقش متغیرهای کمکی در ATS در طی دو مبحث مورد بررسی قرار گیرد. در مبحث اول، نقش متغیرهای کمکی فقط در قالب طراحی نمونه‌گیری مطرح شده و در یک مطالعه موردی از صدف‌های در معرض انقراض آب‌های شیرین مورد بررسی قرار می‌گیرد و در مبحث دوم با فرض معلوم بودن میانگین متغیر کمکی در کل جامعه، نقش این متغیرها علاوه بر مرحله‌ی طراحی نمونه‌گیری، در مرحله‌ی براورد پارامتر مجهول جامعه نیز در قالب براوردگرهای رگرسیونی و نسبتی مورد بررسی قرار می‌گیرد.

۱.۲ نقش متغیرهای کمکی در طراحی نمونه‌گیری در ATS

در این مبحث براساس یک مطالعه موردی از هورنباخ و همکاران [۲۰۱۰] نقش متغیرهای کمکی فقط در مرحله‌ی طراحی نمونه‌گیری مورد بررسی قرار می‌گیرد.
براورد دقیق فراوانی و چگالی جامعه‌های کمیاب معمولا مشکل و پرچالش است. طرح‌های نمونه‌گیری سازوار برای بررسی جامعه‌های کمیاب و خوشه‌ای، توسط تامپسون [۱۹۹۰] مطرح شد و تا کنون کاربردهای زیادی از آن برای مجموعه وسیعی از موقعیت‌های واقعی بسط داده‌شده و مورد بررسی قرار گرفته‌اند. در مقایسه با طرح‌های مرسوم نمونه‌گیری، طرح‌های سازوار کارایی بیشتر (به معنی دقت بیشتر با هزینه ثابت) و قدرت بیشتری در پیدا کردن اعضای کمیاب مدنظر در جامعه دارند (بروان [۲۰۰۳]، اسمیت و همکاران [۲۰۰۴]). کارایی و قدرت کشف اعضای کمیاب به توزیع فضایی و نوع ساختار

خوشه‌ای جامعه مورد نظر بستگی دارد (بروان [۲۰۰۳] ،اسمیت و همکاران [۲۰۰۴]). انتقادی که از طرح‌های نمونه‌گیری خوشه‌ای سازوار می‌شود تصادفی بودن اندازه‌ی نهایی نمونه می‌باشد که می‌تواند بسیار بزرگ و خارج از محدوده توان زمانی و مالی مجریان طرح باشد.
نمونه‌گیری دنباله‌ای دومرحله‌ای (صالحی و اسمیت [۲۰۰۵]، بروان و همکاران [۲۰۰۸]) برای حل این معضل نمونه‌گیری خوشه‌ای سازوار مطرح و بسط داده شد. این نوع از نمونه‌گیری احتیاجی به تعریف همسایگی ندارد و در عین حال همانند طرح نمونه‌گیری خوشه‌ای سازوار توانایی پیدا کردن و کشف جامعه‌های کمیاب را دارد. شرط مورد نظر در این طرح نمونه‌گیری (C)، بطور عمومی براساس یکی از خصیصه‌های متغیر هدف (منظور متغیر اصلی مورد بررسی می‌باشد که در طی این رساله بنابر مقتضیات از آن به عنوان متغیر هدف یا متغیر اصلی نام برده می‌شود)، مانند تعداد اعضای جامعه‌ی هدف در یک واحد نمونه‌گیری تعریف می‌شود.
طرح‌های مختلفی از نوع مرسوم و سازوار موجود هستند که قادر به استفاده از اطلاعات متغیرهای کمکی می‌باشند (تامپسون [۲۰۰۲]). نشان داده شده است که حضور گونه‌های کمیاب، حداقل در بعضی از اجتماعات، با حضور بقیه گونه‌های معمولی موجود در مکان مورد بررسی مرتبط می‌باشند (مه‌یرز و همکاران [۲۰۰۰]، هورنباخ و همکاران [۲۰۱۰]). یافتن و مشاهده اجتماع این نوع از گونه‌های معمولی (غیرکمیاب) بطور معمول ساده تر از گونه‌های کمیاب می‌باشند. حضور و فراوانی این‌گونه از اجتماعات می‌توانند به عنوان متغیر کمکی در طرح نمونه‌گیری برای پیدا کردن متغیرهای کمیاب هدف، مورد استفاده قرار گیرند. بنابراین طرح ATS می‌تواند با تعریف شرط C براساس این گونه از اطلاعات کمکی، از این اطلاعات در طرح نمونه‌گیری بهره جوید.
در ادامه این فصل، نمونه‌گیری دنباله‌ای دومرحله‌ای برای استفاده از متغیرهای کمکی در مرحله‌ی طراحی (همانند تعداد گونه‌های معمولی در یک واحد نمونه‌گیری به عنوان شرط ادامه و تخصیص نمونه‌ی اضافی) بسط داده می‌شود. برای بررسی عملکرد کلی این طرح در استفاده از متغیرهای کمکی، شبیه‌سازی‌هایی با پارامترهای متفاوت براساس مطالعه‌ای برروی صدف‌های در معرض انقراض آبهای شیرین رودخانه مینستوتای آمریکا (هورنباخ و همکاران [۲۰۱۰]) طراحی و مورد استفاده قرار می‌گیرند.

۱.۱.۲ طرح و براوردگر مربوطه

فرض کنید جامعه‌ای N عضوی داریم که در M واحد نمونه‌گیری اولیه (PSU)، هریک شامل N_i واحد نمونه‌گیری ثانویه (SSU) که گاهی از آن‌ها با عنوان "واحد نمونه‌گیری" یاد می‌کنیم، قرار گرفته است. همچنین $\{(i,j), i = 1,2,...,M, j = 1,2,...,N_i\}$ نشان‌دهنده واحد نمونه‌گیری j–ام در i–امین PSU با اندازه یا پاسخ y_{ij} می‌باشد که می‌تواند در کنار این پاسخ یا اندازه‌ی اصلی، یک پاسخ کمکی x_{ij} هم وجود داشته باشد. میانگین متغیر اصلی در i–امین PSU را با $\mu_{yi} = (1/N_i)\sum_{j=1}^{N_i} y_{ij}$ و میانگین کل جامعه برای این متغیر را با $\mu_y = (1/N)\sum_{i=1}^{M} N_i\mu_{yi}$ نمایش می‌دهیم. برای متغیر کمکی

نیز میانگین‌های متناظر بطور مشابه تعریف می‌شوند. در مرحله‌ی اول نمونه‌گیری m تا از PSUها با نمونه‌گیری تصادفی ساده بدون جایگذاری ($SRSWOR$) انتخاب می‌شوند که این نمونه را با s نمایش می‌دهیم. مرحله‌ی دوم طرح شامل دو فاز است، که در فاز اول در تمامی PSUهای انتخاب شده در مرحله‌ی اول، یک نمونه‌ی اولیه s_{1i} به اندازه‌ی n_{1i} گرفته می‌شود و با نشان دادن تعداد واحدهای s_{1i} که در شرط صدق می‌کنند با l_{1i}، در فاز دوم به تعداد $n_{2i} = dg(l_{1i})$ نمونه‌ی اضافی s_{2i} به واحد مربوطه تخصیص داده می‌شود که در آن d یک ثابت از پیش تعیین شده و g یک تابع حقیقی برروی l_{1i} می‌باشد. در آخر تمامی نمونه در i-امین PSU را با s_i نمایش می‌دهیم. قابل توجه است که با تعریف g به صورت $g(l_{1i}) = l_{1i}$ یا $g(l_{1i}) = I\{l_{1i} > 0\}$ به ترتیب یک طرح نمونه‌گیری دنباله‌ای دومرحله‌ای یا نمونه‌گیری دنباله‌ای دومرحله‌ای سازوار خواهیم داشت که در آن I یک تابع نشانگر است. از آن‌جا که در این مبحث قصد نشان دادن کاربرد طرح در بررسی جامعه‌ی صدف‌ها را داریم، می‌توان اشاره کرد که زوج (y_{ij}, x_{ij}) را که به واحد نمونه‌گیری ثانویه j-ام در واحد نمونه‌گیری اولیه‌ی i-ام تخصیص می‌دهیم، نشانگر مقادیر متغیر اصلی (تعداد صدف‌های کمیاب) و متغیر کمکی (تعداد تمامی انواع دیگر صدف‌ها) برای واحد مورد نظر می‌باشد. شرط C می‌تواند به عنوان تابعی از متغیر کمکی، به عنوان مثال به صورت $x_{ij} > c$ یا $(y_{ij} + x_{ij}) > c$، تعریف شود که در آن c یک عدد صحیح مثبت می‌باشد.

تعداد نمونه‌ی اولیه تخصیص داده شده به هر PSU ثابت و برابر n_1 و همچنین n_{2i} برای $l_{1i} > 0$ به صورت ثابت $n_{2i} = n_2$ ($g(l_{1i}) = I\{l_{1i} > 0\}$) در نظر گرفته شدند (البته با $SRSWOR$ در هر دوفاز نمونه‌گیری). با این پارامترها براوردگر مورتی برای i-امین PSU به صورت زیر تعریف می‌شود (صالحی و اسمیت [۲۰۰۵])

$$\hat{\mu}_{yi} = \sum_{j \in s_i} \frac{P(s_i \mid j)}{N_i P(s_i)} y_{ij} \qquad (۱.۲)$$

که در آن $P(s_i)$ و $P(s_i \mid j)$ احتمال گزینش نمونه‌ی s_i و احتمال شرطی گزینش نمونه‌ی s_i به شرط انتخاب واحد j-ام در اولین انتخاب فرایند نمونه‌گیری می‌باشند و بنابراین

$$\frac{P(s_i \mid j)}{P(s_i)} = \begin{cases} \frac{N_i}{n_1}; & \text{for } n_2 = 0 \\ \frac{N_i}{n_1+n_2}; & \text{for } n_2 > 0, l_{1i} > n_2 \\ \frac{N_i(n_1+n_2+1)!}{(n_1+n_2)! - \frac{n_2!(n_1+n_2-l_{1i})!}{n_2-l_{1i}}}; & \text{for } n_2 > 0, l_{1i} \le n_2, y_{ij} \in C \\ \frac{N_i[(n_1+n_2+1)! - n_2!(n_1+n_2-1-l_{1i})/(n_2-l_{1i})!]}{(n_1+n_2)! - \frac{n_2!(n_1+n_2-l_{1i})!}{n_2-l_{1i}}}; & \text{for } n_2 > 0, l_{1i} \le n_2, y_{ij} \in C' \end{cases}$$

که $y_{ij} \in C$ ($y_{ij} \in C'$) بدین معنی است که واحد مورد نظر در شرط C صدق می‌کند (نمی‌کند). همچنین براوردگر هارویتز-تامپسون میانگین کل جامعه (با تعریف $\hat{t}_{yi} = N_i \hat{\mu}_{yi}$) به صورت زیر محاسبه می‌شود

$$\hat{\mu}_y = \frac{1}{N} \sum_{i \in s} \frac{\hat{t}_{yi}}{\pi_i} \qquad (۲.۲)$$

که در آن π_i احتمال انتخاب شدن i-امین PSU در مرحله‌ی اول نمونه‌گیری می‌باشد و با توجه به $SRSWOR$ برابر m/M خواهد بود و واریانس و براوردگر نااریب واریانس آن (در ادامه اثبات می‌شوند) به صورت زیر است

$$var(\hat{\mu}_y) = \frac{1}{N^2}[\frac{M(M-m)}{m}S^2_{t_y} + \frac{M}{m}\sum_{i=1}^{M} var(\hat{t}_{yi})],$$

$$\hat{var}(\hat{\mu}_y) = \frac{1}{N^2}[\frac{M(M-m)}{m}\hat{S}^2_{t_y} + \frac{M}{m}\sum_{i\in s} \hat{var}(\hat{t}_{yi})]$$

و

$$var(\hat{t}_{yi}) = \sum_{j=1}^{N_i}\sum_{j'<j}^{N_i}(1 - \sum_{s_i \ni j,j'} \frac{P(s_i \mid i)P(s_i \mid j')}{P(s_i)})(y_{ij} - y_{ij'})^2$$

و

$$\hat{var}(\hat{t}_{yi}) = \sum_{j=1}^{N_i}\sum_{j'<j}^{N_i}(\frac{P(s_i \mid j,j')}{P(s_i)} - \frac{P(s_i \mid j)P(s_i \mid j')}{P(s_i)})(y_{ij} - y_{ij'})^2$$

که در آن‌ها $P(s_i \mid j,j')$ احتمال گزینش نمونه‌ی s_i به شرط انتخاب واحد j-ام و j'-ام در دو انتخاب اول فرایند نمونه‌گیری می‌باشد و

$$S^2_{t_y} = \frac{1}{M-1}\sum_{i=1}^{M}(t_{yi} - \bar{t}_y)^2,$$
$$\bar{t}_y = \frac{1}{M}\sum_{i=1}^{M} t_{yi},$$
$$\hat{S}^2_{t_y} = \frac{1}{m-1}\sum_{i\in s}(\hat{t}_{yi} - \bar{\hat{t}}_y)^2,$$
$$\bar{\hat{t}}_y = \frac{1}{m}\sum_{i=1}^{m} \hat{t}_{yi}.$$

همچنین تمامی براوردگرهای مطرح شده در این بخش، بطور مشابه برای متغیر کمکی (x) قابل محاسبه می‌باشد.

۲.۱.۲ شبیه‌سازی‌ها

براساس مطالعات هورنباخ و همکاران [۲۰۱۰] یک سری از شبیه‌سازی‌ها طراحی (جدول ۱.۲ را ببینید) و اجرا شده‌اند. در این شبیه‌سازی‌ها متغیر هدف تعداد صدف‌های کمیاب و متغیر کمکی تعداد تمامی انواع صدف‌ها (شامل صدف‌های کمیاب و معمولی) در واحد مربوطه (منظور واحد نمونه‌گیری ثانویه می‌باشد) در نظر گرفته شده‌اند. صدف‌های کمیاب در حال انقراض اغلب به صورت کمیاب و با

ساختار فضایی خوشه‌ای یافت می‌شوند (اسمیت و همکاران [۲۰۰۳] و هورنباخ و همکاران [۲۰۱۰]). هورنباخ و همکاران [۲۰۱۰] چگالی یک نوع از این صدف‌های کمیاب را حدود $0/01m^{-2}$ و کل چگالی صدف‌ها را حدود $1m^{-2}$ تا $3m^{-2}$ گزارش کردند. همبستگی بین متغیر هدف و متغیر کمکی نیز با دو مقدار $0/33$ و $0/67$ گزارش شد. نسبت واحدهای اولیه نمونه‌گیری شده (m/M) بین $0/25$ تا 1 با گام‌های $0/25$ و همچنین نسبت نمونه‌ی اولیه (n_1/N_i) و نسبت نمونه‌ی اضافی تخصیص داده شده (n_2/N_i) بر اساس شرط بین $0/1$ تا $0/4$ با گام $0/1$ در نظر گرفته شدند. برای تعیین شرط C نیز مقدار c از ۰ تا ۵ با گام ۱ در نظر گرفته شد.

برای شبیه‌سازی صدف‌ها از نرم افزار R و بسته $corcounts$ که پارامترهای آن‌ها براساس اطلاعات هورنباخ و همکاران [۲۰۱۰] تعیین شده بود، استفاده کردیم. تعداد خوشه‌ها از توزیع پواسون و مرکز خوشه‌ها بطور تصادفی در منطقه مدنظر تعیین شدند. واحدها در اطراف مرکز خوشه با فاصله تصادفی از توزیع نمایی و در جهتی تصادفی شکل گرفتند (برای اطلاعات بیشتر پیوست (فصل۶) را ببینید).

به‌منظور ارزیابی طرح مدنظر، کارایی طرح و قدرت کشف واحدهای غیر تهی (SSUهایی که شامل حداقل یک صدف کمیاب باشند) محاسبه شده‌اند. کارایی به صورت حاصل تقسیم واریانس میانگین ساده در نمونه‌گیری تصادفی ساده بدون جایگذاری به واریانس براوردگر مورتی در طرح مدنظر محاسبه شد و قدرت کشف نیز به صورت زیر محاسبه شد

$$PP = \frac{p_a}{p_N}$$

که در آن p_N نسبت واحدهای غیرتهی در جامعه و p_a نسبت واحدهای غیرتهی در نمونه‌ی نهایی بدست آمده توسط طرح مدنظر می‌باشند و بزرگتر از ۱ بودن این نسبت نشان از قدرت خوب طرح مدنظر در پیدا کردن اعضای کمیاب نسبت به روش تصادفی ساده (یا طرح‌های مرسوم مشابه) دارد.

چند جامعه شبیه‌سازی شده را در شکل ۱.۲ ببینید.

همچنین برای سنجش میزان اهمیت پارامترهای مهم جمعیتی (مانند چگالی متغیر هدف و متغیر کمکی و همچنین همبستگی بین آن‌ها) و طرح (همانند n_1, n_2, m) از بسته $randomForest$ استفاده شد، که براساس میزان تغییرات در نتایج (از جمله کارایی و قدرت کشف) که بر اثر تغییر یک پارامتر صورت پذیرفته است، پارامترهای مربوطه را از لحاظ اهمیت رتبه‌بندی می‌کند.

۳.۱.۲ نتایج

کارایی برای چگالی متوسط متغیر کمکی ($\mu_x \simeq 1m^{-2}$) و چگالی بسیار پایین متغیر هدف ($\mu_y \simeq 0/01m^{-2}$) همواره، بدون در نظر گرفتن پارامترهای دیگر جامعه و طرح، بزرگتر از ۱ بود (شکل ۲.۲ را ببینید). وقتی که با چگالی متوسط متغیر کمکی، چگالی متغیر هدف پایین بود ($\mu_y \simeq 0/10m^{-2}$) کارایی کمتر از ۱ بود که با افزایش اندازه‌ی نهایی نمونه به بیشتر از ۳۰۰ (نسبت حجمی بیشتر از ۰/۵)،

جدول ۱.۲: هشت جامعه شبیه‌سازی شده با پارامترهای متفاوت چگالی متغیر مدنظر با μ نمایش داده شده است که منظور از چگالی متوسط پاسخ متغیر مدنظر در SSUها می‌باشد. از آن‌جا که در این پژوهش هردو متغیر نشان دهنده تعداد صدف‌ها می‌باشند، چگالی به معنی متوسط تعداد صدف‌ها در هر واحد نمونه‌گیری می‌باشد. CV ضریب تغییرات می‌باشد. ψ نشان دهنده درصد اشغال واحدهای نمونه‌گیری است، که به معنی درصد واحدهای غیرتهی می‌باشد. ρ هم که همان همبستگی بین دو متغیر می‌باشد. همچنین نوشتن امید ریاضی به معنای تولید جامعه‌ای با این خصوصیات توسط نرم‌افزار مربوطه می‌باشد که با توجه به تصادفی بودن فرایند تولید، انتظار می‌رود جامعه‌ای با پارامترهای مدنظر تولید شود.

$\rho(x,y)$	$E(\psi_x)$	$E(CV_x)$	$E(\mu_x)$	$E(\psi_y)$	$E(CV_y)$	$E(\mu_y)$
۰/۳۳	۰/۴	۱/۹۸	۱	۰/۰۱۱	۱۱/۰۴	۰/۰۱
۰/۶۷	۰/۴	۱/۹۸	۱	۰/۰۱۱	۱۱/۰۴	۰/۰۱
۰/۳۳	۰/۷۹	۱/۳۱	۳	۰/۰۱۱	۱۱/۰۴	۰/۰۱
۰/۶۷	۰/۷۹	۱/۳۱	۳	۰/۰۱۱	۱۱/۰۴	۰/۰۱
۰/۳۳	۰/۴	۱/۹۸	۱	۰/۰۸۵	۳/۹۳	۰/۱
۰/۶۷	۰/۴	۱/۹۸	۱	۰/۰۸۵	۳/۹۳	۰/۱
۰/۳۳	۰/۷۹	۱/۳۱	۳	۰/۰۸۵	۳/۹۳	۰/۱
۰/۶۷	۰/۷۹	۱/۳۱	۳	۰/۰۸۵	۳/۹۳	۰/۱

کارایی بیشتر از ۱ می‌شد. همبستگی نیز به نظر می‌رسید که برروی کارایی تاثیر متعادلی (نه خیلی زیاد) داشته باشد. بطوری که همبستگی ۰/۶۷ کارایی نسبتا بیشتری از ۰/۳۳ نشان می‌داد. برای چگالی بالای متغیر کمکی ($\mu_x \simeq 3m^{-2}$) کارایی همواره کوچکتر از ۱ بود.

قدرت کشف طرح مدنظر نیز برای مقدار متوسط چگالی متغیر کمکی $\mu_x \simeq 1m^{-2}$ همواره بزرگتر از ۱ بود. این مقدار برای همبستگی ۰/۶۷ بیشتر از ۰/۳۳ و همچنین برای چگالی متغیر هدف ۰/۰۱ بیشتر از ۰/۱۰ بود. برای مقدار بالای چگالی متغیر کمکی (۳) حدود ۱ یا کمی بیشتر از مقدار ۱ برای اندازه بالای نمونه بود (شکل ۲.۲ را ببینید).

برای اهمیت پارامترهای جمعیتی و طرح، بر طبق جدول ۲.۲ (که در آن عدد بزرگتر نشانگر اهمیت بیشتر است) چگالی متغیر کمکی بیشترین اهمیت را از نظر هردو معیار کارایی و قدرت کشف داشت. چگالی متغیر هدف و تعداد PSUهای انتخاب شده (m) در درجه دوم اهمیت از نظر کارایی بودند، در صورتی که از نظر قدرت کشف، تعداد نمونه‌ی اضافی تخصیص داده شده در فاز دوم (n_2) مهم‌تر بود.

۴.۱.۲ جمع بندی و پیشنهادات

همان‌طور که می‌دانیم طرح‌های نمونه‌گیری زیادی وجود دارند که قادر به استفاده از متغیرها و اطلاعات کمکی در بهبود طرح (از نظر کارایی و ...) می‌باشند. از آن‌جا که این طرح‌ها منوط به