دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

برآورد ناحیه کوچک با استفاده از مدل‌های خطی و خطی تعمیم یافته با اثرهای آمیخته

پایان‌نامه کارشناسی ارشد (آمار اقتصادی و اجتماعی)

پیام مختاریان دهکردی

اساتید راهنمای پایان‌نامه

دکتر امیر نادری

دکتر محمد صالحی مرزیجرانی

۱۳۸۸

دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد (آمار اقتصادی و اجتماعی) آقای پیام مختاریان دهکردی

تحت عنوان

برآورد ناحیه کوچک با استفاده از مدل‌های خطی و خطی تعمیم یافته با اثرهای آمیخته

در تاریخ ۱۳۸۸/۳/۳۱ توسط کمیته تخصصی زیر مورد بررسی و تصویب نهائی قرار گرفت.

۱— استاد راهنمای پایان‌نامه	دکتر امیر نادری
۲— استاد راهنمای پایان‌نامه	دکتر محمد صالحی مرزیجرانی
۳— استاد داور ۱	دکتر هوشنگ طالبی (دانشگاه اصفهان)
۴— استاد داور ۲	دکتر ایرج کاظمی

سرپرست تحصیلات تکمیلی دانشکده دکتر رسول نصر اصفهانی

با تشکر از پدر و مادر عزیزم و همسرم
که مرا در طول زندگی و تحصیلات
یاری دادند.

فهرست مطالب

چکیده:

تکنیک برآورد ناحیه کوچک به طور عمده متکی بر مدل‌های آمیخته با اثرهای تصادفی ناحیه‌ای می‌باشد. بر این مبنی با استفاده از روش‌های متداول در برآورد پارامترها در مدل‌های آمیخته به دنبال یک برآورد برای ناحیه کوچک هستیم و در این میان به معرفی بهترین پیش‌بینی کننده‌ی خطی نااریب (BLUP) برای پارامتر مورد نظر ناحیه کوچک می‌پردازیم. اما در مقابل این رویکرد، رویکرد دیگری مبتنی بر توابع حساسیت و رگرسیون M—چندک معرفی شده است. این رویکرد نیازمند مفروضات متعارفی چون نرمال بودن و مسائل مربوط به توصیفات اثرهای تصادفی نمی‌باشد. از آن جایی که این مفروضات و همچنین وجود نقاط پرت مشکلاتی در برآورد ایجاد می‌کنند و نتیجتاً برآوردی نامناسب (نااریب و یا ناکارا) به دست می‌آید این رویکرد را به عنوان هدف اصلی این پایان‌نامه دنبال می‌کنیم. در این پایان‌نامه سعی بر آن است که علاوه بر معرفی کاملی از مدل‌هایی با اثرهای آمیخته و نحوه‌ی برآورد ناحیه کوچک با استفاده از این مدل‌ها به رویکرد نیرومند برای برآورد ناحیه کوچک مبتنی بر دو روش استفاده از توابع تاثیر کراندار و رگرسیون M—چندک بپردازیم.

طبقه‌بندی موضوعی: ۶۲D۰۵، ۶۲J۱۲، ۶۲G۳۵

کلید واژگان: برآورد دامنه، برآورد ناحیه کوچک، بهترین پیش‌بینی کننده‌ی نااریب خطی، تابع تاثیر کراندار، مدل‌های آمیخته، نیرومندی

فصل ۱

مقدمه

در به دست آوردن اطلاعات از یک متغیر یا مجموعه‌ای از متغیرها بررسی‌های نمونه‌ای کاربرد بسیاری دارند. در بررسی‌های نمونه‌ای یا به طور کلی روش‌شناسی آمارگیری هدف تنهابه دست آوردن اطلاعات یا برآورد پارامترهای جامعه نیست بلکه در بسیاری از موارد برآورد به دست آوردن پارامترها در زیرجامعه‌هایی از جامعه‌ی مورد نظر می‌باشد که اصطلاحاً این زیر جامعه‌ها را دامنه می‌نامیم. دامنه را ممکن است به صورت نواحی جغرافیایی یا گروه‌های آمارگیری اجتماعی[۱] و یا زیرجامعه‌هایی در جوامع بزرگ تعریف کنیم. به عنوان مثال ایالت‌ها، استان‌ها، شهرستان‌هاو نواحی شهرها نمونه‌هایی از نواحی جغرافیایی می‌باشند و یا این که گروه‌های مختلف سنی، نژادی و جنسی درون نواحی جغرافیایی می‌توانند به عنوان گروه‌های آمارگیری اجتماعی در نظر گرفته شوند. برای هر جامعه می‌توانیم بر حسب نیاز و شرایط زیرجامعه (دامنه) تعریف کنیم.

در بررسی‌های، نمونه‌ای روش‌های مختلفی برای برآورد کردن دامنه وجود دارد که در این پایان‌نامه به معرفی این روش‌ها می‌پردازیم. یک برآورد برای دامنه را برآورد مستقیم می‌نامیم هر گاه این برآورد تنها مبتنی بر داده‌های درون دامنه باشد. در برآورد دامنه می‌توانیم از یک متغیر کمکی معلوم درون آن دامنه برای برآورد کردن و بهبود بخشیدن به کارایی برآوردگر استفاده کنیم؛ مانند کل یک متغیر کمکی X که در ارتباط با متغیر مورد نظر y می‌باشد. به طور کلی برآوردگرهای مستقیم را طرح—پایه می‌نامیم. یکی دیگر از روش‌های برآورد دامنه، روش غیر مستقیم می‌باشد. در این روش علاوه بر استفاده از داده‌های درون

[۱] Socio-Demographic

ناحیه از سایر داده‌ها و اطلاعات درون دامنه‌های دیگر نیز استفاده می‌کنیم. با استفاده از مدل سازی میان اطلاعات درون دامنه با سایر دامنه‌ها می‌توانیم برآوردگری مناسب برای پارامترهای مورد نظر به دست آوریم. در مواردی که با حجم نمونه‌ی کم درون ناحیه برخورد می‌کنیم برآوردگرهای مدل–پایه انتخاب مناسبی می‌باشند. در فصل دوم این پایان‌نامه به معرفی و توصیف تعدادی از این برآوردرگرها می‌پردازیم.
یک دامنه یا ناحیه را بزرگ می‌نامیم اگر حجم نمونه‌ی درون آن ناحیه زیاد باشد و آن ناحیه را کوچک می‌نامیم که حجم درون ناحیه به اندازه‌ی کافی زیاد نباشد. بر اساس این دو نوع تقسیم بندی در نواحی آمارگیری، برآوردگرهای مختلفی ارائه شده است که آن‌ها در این نوع تقسیم بندی با یک دیگر متفاوت هستند و این باعث ایجاد یک مبحث جدید و گسترده در زمینه‌ی روش‌شناسی آمارگیری به نام برآورد ناحیه کوچک شده است. در این مبحث از عباراتی چون نواحی محلی[۲]، زیردامنه[۳]، زیرگروه کوچک[۴] یا دامنه‌ی جزئی[۵] نیز به جای ناحیه کوچک استفاده می‌شود.
در بسیاری از کاربردهای عملی ممکن است درون دامنه‌ی مورد نظر نمونه وجود نداشته باشد یا این که حجم نمونه بسیار کم باشد و بنابراین تنها راه ممکن برای برآورد در دامنه استفاده از اطلاعات و داده‌های مرتبط و متناظر درون دامنه‌ای دیگر بر اساس روش‌های مدل–پایه است. از آن جا که در مبحث برآورد ناحیه کوچک تمام کار بر اساس مدل سازی می‌باشد باید به طور عمده‌ای به این مسئله و روش‌های مختلف برآورد در آن به پردازیم.
همان طور که اشاره شد برآوردگرهای غیر مستقیم مبتنی بر مدل‌های ناحیه کوچک را برآوردگرهای مدل–پایه می‌نامیم. مدل‌های ناحیه کوچک را به دو دسته‌ی عمده تقسیم می‌کنیم:
۱) مدل‌های سطح ناحیه‌ای که در آن‌ها برآوردگرهای مستقیم ناحیه کوچک مرتبط با متغیرهای کمکی توصیف کننده‌ی ناحیه می‌باشند.
۲) مدل‌های سطح واحدی که در آن‌ها مقادیر واحدهای متغیر مورد بررسی مرتبط با متغیرهای کمکی توصیف کننده‌ی واحد می‌باشند.
در فصل سوم به طور کامل به معرفی این دو دسته از مدل‌های ناحیه کوچک می‌پردازیم. در انتهای همین فصل نیز به معرفی دسته‌ای دیگر از مدل‌های مورد استفاده در برآورد ناحیه کوچک اشاره می‌کنیم. برای متغیرهای پاسخ، y، به صورت پیوسته از مدل‌های اثر آمیخته خطی کلی و برای حالتی که متغیر y به صورت رسته‌ای یا دو دویی است از مدل‌های اثر آمیخته خطی تعمیم یافته استفاده می‌کنیم. توجه این پایان‌نامه به برآوردگرهای مدل–پایه‌ی ناحیه کوچک بر اساس مدل‌های اثر آمیخته می‌باشد لذا به ترتیب

[۲] Local Area
[۳] Subdomain
[۴] Small Subgroup
[۵] Minor Domain

در فصل‌های چهارم و پنجم به طور مشروح به معرفی و خواص مدل‌های اثر آمیخته خطی و خطی تعمیم یافته می‌پردازیم. بر اساس این دو فصل، زمینه‌ای برای چگونگی استفاده از این مدل‌ها را در برآورد ناحیه کوچک ایجاد می‌کنیم.

در فصل ششم با استفاده از نتایج به دست آمده در فصل‌های قبل به برآورد و معرفی یک پیش‌بینی کننده برای میانگین و یا کل دامنه می‌پردازیم. در برآورد ناحیه کوچک سعی می‌شود یک پیش‌بینی کننده‌ی نااریب خطی برای پارامتر مورد نظر پیدا کنیم. در این میان پیش‌بینی کننده‌های نااریب خطی زیادی به دست می‌آیند از این رو به دنبال پیش‌بینی کننده‌ای می‌گردیم که در میان سایر پیش‌بینی کننده‌ها از واریانس (MSE) کمتری برخوردار باشد. پیش‌بینی کنندی به دست آمده را بهترین پیش‌بینی کننده‌ی نااریب خطی[۶] (BLUP) می‌نامیم. از آن جا که در این میان با مسئله‌ی برآورد مولفه‌های واریانس، که عمدتاً نا معلوم هستند، برخورد می‌کنیم با برآورد کردن این مولفه‌ها پیش‌بینی کننده، به بهترین پیش‌بینی کننده‌ی نااریب خطی تجربی[۷] (EBLUP) تبدیل می‌شود. بر اساس روابط بین نواحی و مدل‌های اثر آمیخته خطی، بهترین پیش‌بینی کننده‌ی نااریب خطی تجربی (EBLUP) را برای پارامتر مورد نظر در ناحیه کوچک به دست می‌آوریم. در این فصل به این موضوع به طور کامل می‌پردازیم.

یکی از مشکلاتی که در برآورد کردن به ویژه در برآوردهای مدل–پایه وجود دارد بر قرار نبودن فرض‌های اولیه‌ی مدل سازی (به ویژه فرض نرمال بودن توزیع خطاها) و وجود نقاط پرت می‌باشد. برآوردهای ناحیه کوچک هم از وجود این مشکل مستثنی نیستند. یکی از مشکلات عمده در برآوردهای ناحیه کوچک وجود نقاط پرت است. از آن جا که در این زمینه ممکن است حجم نمونه درون ناحیه بسیار کم باشد، وجود نقاط پرت باعث تاثیر زیاد بر روی برآوردگر و برآوردگری با کارایی کم خواهد بود. از این رو باید از روشی استفاده کنیم که کنترلی بر این مسئله داشته باشیم. استفاده از روش‌های برآورد نیرومند گزینه‌ای مناسب برای فایق آمدن بر این مشکل می‌باشد. در برآوردهای نیرومند یکی از روش‌هایی که اثر نقاط پرت را کنترل می‌کند استفاده از توابع تاثیر کراندار[۸] می‌باشد. در فصل هفتم به تعریف و بررسی برآورد نیرومند ناحیه کوچک مبتنی بر توابع تاثیر کراندار می‌پردازیم. همچنین با استفاده از رگرسیون M–چندک[۹] برآوردی نیرومند برای توزیع و میانگین ناحیه کوچک ارائه می‌دهیم و در انتهای همین فصل یک رویکرد نیرومند در برآورد ناحیه کوچک، با استفاده از مدل‌های اثر آمیخته خطی تعمیم یافته‌ی سلسله مراتبی، معرفی می‌کنیم.

برآورد ناحیه کوچک یکی از مباحث پرکاربرد در روش‌شناسی آمارگیری می‌باشد که بیشتر در حوزه‌ی

---

[۶] Best Linear Unbiased Predictor

[۷] Empirical Best Linear Unbiased Predictor

[۸] Bounded Influence Function

[۹] M-Quantile

علوم اجتماعی و اقتصادی به کار می‌رود. در ادامه به بیان چند مثال معروف در این رابطه می‌پردازیم.

برآورد درآمد سرانه برای نواحی کوچک

این مثال بیان‌گر کاربرد EBLUP مبتنی بر یک مدل سطح ناحیه‌ای در برآورد درآمد درون نواحی کوچک است. اصل این روش توسط فی و هریوت[۱۰] (۱۹۷۹) پیشنهاد شد. بر اساس روش پیشنهادی آن‌ها کمیته‌ی سرشماری آمریکا[۱۱] (USCB) برآوردی برای درآمد سرانه (PCI) در نواحی کوچک با استفاده از اطلاعات به دست آمده از سرشماری سال ۱۹۹۰ و همچنین داده‌های مربوط به مالیات‌ها و درآمدهای نواحی بزرگ‌تر به دست آورد و با یک مدل سازی مناسب بر روی این اطلاعات، برآوردی دقیق و کارا برای درآمد سرانه در نواحی کوچک ارائه گردید.

میزان فقر

روش پیشنهادی فی و هریوت نیز در برآورد فقر در میان دانش آموزان آمریکا توسط هیات تحقیقات ملی[۱۲] (NRC) در سال ۲۰۰۰ به کار گرفته شد. آن‌ها با استفاده از اطلاعات ثبتی پارلمان آموزشی درون ایالتی مبنی بر وضعیت درآمد سرانه‌ی خانواده‌ی دانش آموزان، تعداد فرزندان خانواده‌ی آن‌ها و وضعیت بهداشت و تغذیه‌ی آن‌ها و همچنین اطلاعات به دست آمده از سرشماری قبلی، مدلی ارتباط دهنده بین پارامتر میزان فقر و داده‌های به دست آمده پیشنهاد دادند و بر آن اساس پیش‌بینی کننده‌ای برای میزان فقر در میان دانش آموزان آمریکا به دست آوردند. شایان ذکر است که مدارس آمریکا به عنوان ناحیه کوچک در نظر گرفته شده است.

کشاورزی

سازمان کشاورزی ملی آمریکا[۱۳] (USNAO) در سال ۲۰۰۱ در صدد برآورد سطوح زیر کشت گندم و محصول به دست آمده‌ی آن در ایالت کانزاس[۱۴] برآمد. این سازمان با استفاده از داده‌های به دست آمده از عکس برداری‌های ماهواره‌ای به عنوان متغیر کمکی، و همچنین میزان مالیات دریافتی از زارعان آن ایالت مدلی ارتباط دهنده با میزان محصول درو شده ارائه کرد. بر اساس آن مدل، پیش‌بینی کننده‌ای برای میزان محصول درو شده و به طبع برآوردی برای مساحت زمین‌های زیر کشت گندم در آن ایالت به دست آورد. زمین‌های تحت کشت گندم در آن ایالت به عنوان نواحی کوچک در نظر گرفته شده است.

[۱۰] Fay & Herriot
[۱۱] U.S. Census Bureau
[۱۲] National Research Council
[۱۳] U.S. National Agricultural Organization
[۱۴] Kansas

در پیوست این پایان‌نامه تعریف‌ها و قضایای مورد نیاز فصل‌ها را به اختصار آورده‌ایم. در پیوست (۱) بسته‌ی بارگذاری شده‌ی SAE2 بر روی نرم‌افزار R را با ذکر یک مثال معرفی می‌کنیم.

فصل ۲

برآورد دامنه

پیشگفتار

داده‌های نمونه‌ای به صورت گسترده‌ای نیاز به یک برآورد دقیق و قابل اطمینانی از کل و یا میانگینی از نواحی بزرگ و یا دامنه‌ها دارند. یک برآوردگر مستقیم برای یک دامنه می‌بایستی که از مقادیر در دسترس متغییر مورد نظر ما y باشد که این متغییر باید از درون واحدهای نمونه‌گیری در دامنه به دست آمده باشد. روش‌های مبتنی بر مدل سازی یکی از روش‌های مورد استفاده برای برآورد مستقیم و استنباط درباره‌ی آن‌ها می‌باشد. در این بخش به معرفی و توضیح درباره‌ی برخی از روش‌های برآورد مستقیم دامنه می‌پردازیم. در صورت وجود حجم نمونه‌ی کافی (بزرگ) برای برآورد دامنه از برآورد مستقیم استفاده می‌کنیم. اما در برآوردهای ناحیه کوچک معمولاً نواحی کوچکی که انتخاب شده‌اند فاقد واحد نمونه گیری هستند و یا این که دارای حجم نمونه‌ی بسیار کمی می‌باشند، لذا از این رو وجود این مشکل باعث می‌شود که برای برآورد دامنه از برآوردهای غیر مستقیم استفاده کنیم که در بخش دوم این فصل به معرفی چند نمونه از برآوردهای غیر مستقیم می‌پردازیم.

۲—۱ برآورد مستقیم دامنه

۲—۱—۱ برآورد بدون اطلاعات کمکی

فرض می‌کنیم U_i بیانگر دامنه‌ی (زیر جامعه‌ی) مورد نظر ما باشد و ما به دنبال برآورد کل دامنه، $Y_i = \sum_{U_i} y_i$ یا میانگین دامنه، $\bar{Y}_i = Y_i/N_i$ باشیم که در آن N_i تعداد اعضای درون واحد U_i است. در صورتی که y_i دو دوئی باشد (صفر و یک) آن گاه $\bar{Y}_i$ را به صورت نسبت دامنه، P_i، در نظر می‌گیریم. یک عملگر بر روی Y به نام $Y(y)$ به صورت زیر تعریف می‌کنیم:

$$y_{ij} = \begin{cases} y_i & if\ j \in U_i \\ ۰ & otherwise \end{cases}$$

و

$$a_{ij} = \begin{cases} ۱ & if\ j \in U_i \\ ۰ & otherwise \end{cases}$$

بنابراین داریم:

$$Y(y_i) = \sum_{j \in U} y_{ij} = \sum_{j \in u_i} y_i = Y_i \qquad (۱)$$

و

$$Y(a_i) = \sum_{j \in U} a_{ij} = \sum_{j \in u_i} ۱ = N_i \qquad (۲)$$

از این رو y_{ij} را می‌توانیم به صورت $a_{ij}y_{ij}$ بنویسیم. حال می خواهیم از یک نمونه درون دامنه به برآورد آن دامنه به پردازیم در شرایطی که اطلاعات کمکی از جامعه در دست نداریم. بر اساس تعریف فوق برآوردگر را به صورت

$$\hat{Y}_i = \hat{Y}(y_i) = \sum_{j \in s} w_{ij} y_{ij} = \sum_{j \in s_i} w_j y_j \qquad (۳)$$

می‌باشد که در آن S واحدهای نمونه گیری و S_i واحد نمونه گیری از دامنه‌ی i—ام و $w_j = w_j(s) = \frac{۱}{\pi_j}$ می‌باشد، همچنین $\pi_j = \sum_{\{s:j\in s\}} p(s)$ و $j \in s$. در واقع π_j احتمال انتخاب عضو j—ام در نمونه‌گیری از واحد S_i می‌باشد. از طرفی واریانس این برآوردگر به صورت

$$V(\hat{Y}) = -\sum_{j<k}\sum_{j,k\in s} w_{jk}(s)\pi_j\pi_k\left(\frac{y_j}{\pi_j} - \frac{y_k}{\pi_k}\right)^{۲}$$

است و در آن $w_{jk} = w_{jk}(s) = \frac{\pi_{jk}}{\pi_j\pi_k} - ۱$ که π_{jk} احتمال انتخاب توام عضوهای j—ام و k—ام در نمونه می‌باشد. برای مطالعه‌ی بیشتر در این رابطه را می‌توان به سارندل[۱] و همکاران (۱۹۹۲) مراجعه داد. بنابر تعریف می‌بینیم که میانگین دامنه برابر است با $\bar{Y}_i = Y(y_i)/Y(a_i)$ و لذا برآورد آن به صورت زیر است:

$$\hat{\bar{Y}}_i = \frac{\hat{Y}(y_i)}{\hat{Y}(a_i)} = \frac{\hat{Y}_i}{\hat{N}_i} \qquad (۴)$$

۲—۱—۲ برآورد با وجود اطلاعات کمکی

گاهی اوقات متغییر کمکی X را که دارای یک رابطه‌ی آماری معنی داری با متغییر مورد نظر یعنی y است را در اختیار داریم. بر اساس وجود این رابطه می‌توانیم یک برآوردگر نااریب برای Y ارائه دهیم. برای بیان این موضوع ابتدا اشاره به برآورد رگرسیونی تعمیم یافته[۲] (GREG) می‌کنیم.

۲—۱—۳ برآورد رگرسیونی تعمیم یافته

فرض می‌کنیم که $X = (X_۱, \ldots, X_p)^T$ اطلاعات کمکی (متغییر کمکی) در کل دامنه (جامعه) باشد که بردار x_{ij} مشاهده‌ی j—ام $(j \in s)$ از X است. داده‌های ما به صورت (y_j, X_j)، $j \in s$، می‌باشند. بر اساس این داده‌ها برآوردگر GREG را به صورت زیر تعریف می‌کنیم:

$$\hat{Y}_{GR} = \hat{Y} + (X - \hat{X})^T\hat{B} \qquad (۵)$$

که $\hat{X} = \sum_s w_j X_j = \hat{Y}(X)$ و $\hat{B} = (\hat{B}_۱, \ldots, \hat{B}_p)^T = \hat{B}(y)$ می باشد. $\hat{B}(y)$ را می‌توانیم از حل معادله‌ی کمترین مربعات وزنی، به صورت زیر به دست آوریم:

$$\Big(\sum_s w_j X_j {X_j}^T/c_j\Big)\hat{B} = \sum_s w_j X_j y_j/c_j \qquad (۶)$$

---

[۱] Sarndal et al.

[۲] Generalized Regression Estimator

در این جا $c_j > ۰$ را به عنوان یک ثابت معین[۳] معرفی می کنیم. فرم بسته‌ی $\hat{Y}_{GR}$ را به صورت برآوردگر

$$\hat{Y}_{GR} = \sum_s w_j{}^* y_j = \hat{Y}_{GR}(y) \qquad (۷)$$

که $w_j{}^* = w_j{}^*(s) = w_j(s) g_j(s)$ و $g_j(s) = ۱ + (X - \hat{X})^T (\sum_s w_j X_j X_j{}^T)^{-۱} X_j / c_j$، برای برآورد درون یک دامنه می‌نویسیم. این حالت کاملاً شبیه حالت کلی است با این تفاوت که نمادها اندکی متفاوت هستند. در این صورت برآوردگر GREG برای Y_i به صورت زیر نوشته می‌شود:

$$\hat{Y}_{iGR} = \hat{Y}_{GR}(y_i) = \sum_{j \in s_i} w_j{}^* y_j \qquad (۸)$$

که در آن $y_j = a_{ij} y_j$. یک حالت خاص از این برآوردگر همان برآوردگر نسبتی است که به صورت زیر می‌باشد:

$$\hat{Y}_{iR} = \hat{Y}_R(y_i) = \frac{\hat{Y}_i}{\hat{X}} X \qquad (۹)$$

برآوردگر GREG دارای خواص زیر است:

- $\sum_{i=۱}^r \hat{Y}_{GR}(y_i) = \hat{Y}_{GR}(\sum_{i=۱}^r y_i)$
- برآوردگر $\hat{Y}_{iGR}$ به طور تقریبی در صورتی که حجم نمونه زیاد باشد نااریب است.
- در صورتی که حجم نمونه‌ی درون دامنه زیاد باشد این برآوردگر سازگار است.

برای اثبات موارد فوق می‌توان به ککران[۴] (۱۹۷۷) مراجعه کرد.

۲–۱–۴ برآوردگرهای مستقیم بهبود یافته

در این جا به بیان برآوردگرهای مستقیم بهبود یافته[۵] می‌پردازیم که در آن مقادیر y خارج از دامنه‌ی مورد نظر ماست. با توجه به این که می‌خواهیم نااریبی حفظ شود با جایگذاری ضریب رگرسیونی B_i به جای

[۳] Specified Constant
[۴] Cochran
[۵] Modified Direct Estimator

$\hat{B}_i$ در رابطه‌های قبل، برآوردگر را به صورت زیر می‌نویسیم:

$$\tilde{Y}_{iGR} = \hat{Y}_i + (X_i - \hat{X}_i)^T \hat{B} = \sum_{j \in s} \tilde{w}_{ij} y_j \qquad (۱۰)$$

که $\tilde{w}_{ij} = w_j a_{ij} + (X_i - \hat{X}_i)(\sum_s w_j X_j X_j{}^T / c_j)^{-1}(w_j X_j / c_j)$ و a_{ij} متغییر نشانگر دامنه می‌باشند. در این حالت برآوردگر $\tilde{Y}_{iRG}$ نااریب است حتی اگر حجم نمونه کوچک باشد. حالت خاصی از این برآوردگر چنین است:

$$\tilde{y}_{iR} = \hat{Y}_i + \frac{\hat{Y}}{\hat{X}}(X_i - \hat{X}_i) \qquad (۱۱)$$

این برآوردگر، برآوردگر رگرسیونی آمارگیری[۶] (سرشماری) نیز نامیده می‌شود. برای مطالعه‌ی بیشتر در مورد این برآوردگر می‌توان به بتیس[۷] و همکاران (۱۹۸۸) و وودروف[۸] (۱۹۶۶) مراجعه نمود.

در این بخش به اختصار به معرفی برخی از برآوردگرهای مستقیم معمول برای برآورد کل یا میانگین دامنه پرداختیم. هدف و توجه اصلی ما در این پایان‌نامه بر روی برآوردگرهای غیر مستقیم مدل–پایه است که در ادامه به تفصیل به آن می‌پردازیم.

۲–۲ برآورد غیر مستقیم دامنه

۲–۲–۱ برآورد ترکیبی

گاهی اوقات تعدادی از نواحی کوچک ما دارای مشخصه‌های یکسانی هستند که مجموعه‌ی این نواحی کوچک ناحیه‌ای بزرگ را ایجاد می‌کند. در این شرایط می‌توانیم با استفاده از یک برآوردگر مستقیم از نواحی بزرگ، برآوردگری غیر مستقیم از نواحی کوچک به دست آوریم که اصطلاحاً به این نوع برآوردگرها، برآوردگر ترکیبی[۹] (سینتتیک) می‌گویند.

---

[۶] Survey Regrassion Estimator
[۷] Battese
[۸] Woodruff
[۹] Synthetic Estimator