جمهوری اسلامی ایران
مرکز آمار ایران
پژوهشکده‌ی آمار

برآورد خطای نمونه‌گیری در آمارگیری‌های نمونه‌ای پیچیده

تابستان ۱۳۸۴

برآورد خطای نمونه‌گیری در آمارگیری‌های نمونه‌ای پیچیده

صدیقه شمس
فرشید جمشیدی
زهره فلاح‌محسن‌خانی
فاطمه هرندی
عباس گرامی

پژوهشکده آمار
تابستان ۸۴

به نام خداوند جان و خرد

پیشگفتار

خطای نمونه‌گیری عبارت است از تغییرپذیری تصادفی که به دلیل بررسی نمونه‌ای از جامعه به جای تمام واحدهای جامعه، رخ می‌دهد. زمانی که کاربران از داده‌های آمارگیری‌های نمونه‌ای در برنامه‌ریزی استفاده می‌نمایند، دانستن خطای نمونه‌گیری امری مهم است. گزارش این خطا هم زمان با ارائه‌ی برآوردها، به برنامه‌ریزان این اجازه را می‌دهد که درتصمیم‌گیری‌ها آگاهانه‌تر عمل نماید. محاسبه و ارائه‌ی منظم این اطلاعات، اهداف دیگری هم‌چون ارزیابی و اصلاح طرح‌های آمارگیری را نیز پوشش می‌دهد.

طرح‌های نمونه‌گیری که در تولید آمارهای رسمی مورد استفاده قرار می‌گیرد، اغلب طرح‌های پیچیده‌ای هستند که برای بهینه‌سازی طرح و افزایش دقت نتایج، مبتنی بر نمونه‌گیری خوشه‌ای، طبقه‌بندی، روش‌های برآورد نسبتی و یا رگرسیون و ... می‌باشند. در این موارد به دلیل پیچیدگی روش‌های مستقیم محاسبه‌ی خطای نمونه‌گیری و در بعضی موارد عدم امکان محاسبه‌ی خطای نمونه‌گیری به روش مستقیم، لازم است از روش‌های غیر مستقیم برای این منظور استفاده شود.

در این راستا پژوهشکده‌ی آمار با تشکیل یک گروه مطالعاتی، طرح پژوهشی «خطای نمونه‌گیری در آمارگیری‌های نمونه‌ای پیچیده» را در دستور کار خود قرار داد. این گروه روش‌های غیر مستقیم برآورد واریانس مانند خطی‌سازی، تکرار متعادل، گروه‌های تصادفی، جک‌نایف و بوت‌استرپ را به تفصیل بررسی نمود و با استفاده از چند سری داده‌ی شبیه‌سازی شده که هر یک ۱۰۰ بار تکرار شده بودند، مقایسه‌ای بین این روش‌ها انجام داد. در انتها نیز خطای نمونه‌گیری برآورد فصل زمستان طرح اشتغال و بیکاری سال ۱۳۸۲ مرکز آمار ایران را برآورد نمود.

در گروه مطالعاتی طرح مذکور، خانم دکتر صدیقه شمس به عنوان مجری طرح و آقایان فرشید جمشیدی و دکتر عباس گرامی و خانم‌ها فاطمه هرندی و زهره فلاح‌محسن‌خانی به عنوان همکار عضویت داشتند که بدین وسیله از زحمات یکایک این افراد تشکر و قدردانی می‌شود.

در طول انجام این طرح، خانم‌ها افسانه زنگنه و مهرنوش میرمحمد نیز زحمت تایپ مستندات طرح را به عهده داشتند که در این جا لازم است از تلاش این عزیزان تشکر و قدردانی شود.

اگـرچه در انجـام پـژوهش، داور محـترم نظـرات اصـلاحی خود را اعلام نموده و از این نظرات در جهـت بهبود هر چه بیش‌تر طرح استفاده شده است، اما از خوانندگان محترم تقاضا دارد موارد احتمالی اشـکالات و ابهامـات موجـود در نشـریه را بـه گـروه پژوهشـی طـرح‌هـای فنـی و روش‌هـای آماری پژوهشکده‌ی آمار منعکس نمایند.

گروه پژوهشی طرح‌های فنی و روش‌های آماری

پژوهشکده آمار

فهرست مطالب

عنوان **صفحه**

بخش اول: کلیات

۱-مفهوم خطای نمونه‌گیری **۳**

۱-۱- واریانس نمونه‌گیری ۶

۱-۲- اهمیت آگاهی از خطای نمونه‌گیری ۸

۱-۳- جنبه‌های عملی در انتخاب روش محاسبه‌ی خطای نمونه‌گیری ۹

۲- مفهوم آمارگیری پیچیده **۱۰**

بخش دوم: روش شناسی

۳-روشی ساده برای تقریب واریانس **۱۷**

۳-۱- تعمیم روش به سایر آماره‌های پیچیده ۲۰

۳-۲- کاربردهای خاص ۲۲

۳-۳- کاربرد روش در عمل ۲۴

۴- روش توابع واریانس تعمیم یافته **۲۶**

۴-۱- مدل‌های واریانس تعمیم یافته ۲۷

۴-۲- نحوه محاسبه GVF ۲۸

۵-روش خطی سازی **۲۹**

۵-۱- تقریب‌های خطی در جامعه نامتناهی ۳۰

۵-۲- تقریب‌های خطی در جامعه متناهی ۳۲

۵-۳- یک الگوریتم محاسباتی ۳۶

۵-۴- چند مثال ۳۸

۶-روش گروه‌های تصادفی **۴۲**

۶-۱- گروه‌های تصادفی مستقل ۴۲

۶-۲- گروه‌های تصادفی وابسته ۴۴

۶-۲-۱- یک فرایند عمومی برآورد ۴۶

۶-۲-۲- برآوردگرهای خطی با نمونه‌گیری تصادفی از واحدهای اولیه ۴۷

۶-۲-۳- برآوردگرهای خطی با نمونه‌های خوشه‌بندی شده ۵۰

۶-۲-۴- روش عمومی برآورد واریانس برای نمونه‌گیری بدون جایگذاری ۵۳
۶-۳- پایداری برآوردگرهای گروه تصادفی ۵۵
۶-۴- انحرافات از اصول مشخص ۵۶

۷-روش‌های بازنمونه‌گیری در برآورد واریانس برآوردگرها ۵۸

۷-۱- روش جک‌نایف ۵۸
۷-۱-۱- روش جک‌نایف در جوامع نامتناهی ۵۹
۷-۱-۲- برآورد مقدار اریبی با استفاده از روش جک‌نایف ۶۱
۷-۱-۳- جک‌نایف تعمیم یافته ۶۲
۷-۱-۴- روش جک‌نایف با حذف d مشاهده در هر بار ۶۳
۷-۱-۵- روش جک‌نایف در آمارگیری‌های نمونه‌ای تصادفی ساده ۶۵
۷-۱-۶- استفاده از روش جک‌نایف در نمونه‌گیری با طبقه‌بندی ۶۷
۷-۱-۷- استفاده از روش جک‌نایف در نمونه‌گیری خوشه‌ای ۷۲
۷-۱-۸- استفاده از روش جک‌نایف در نمونه‌گیری طبقه‌بندی خوشه‌ای ۷۳
۷-۲- روش تکرار مکرر متعادل (BRR) یا نیم نمونه‌ی متعادل (BHS) ۷۴
۷-۲-۱- کاربرد در طرح‌های چند مرحله‌ای ۷۸
۷-۲-۲- کاربرد در برآوردگرهای غیر خطی ۷۹
۷-۲-۳- نمونه‌گیری بدون جایگذاری ۸۲
۷-۲-۴- متعادل‌سازی جزیی ۸۳
۷-۲-۵- تعمیم روش BRR در حالت $n_h \neq 2$ ۸۶
۷-۳- روش بوت‌استرپ (خودگردان) ۸۹
۷-۳-۱- روش بوت‌استرپ ساده ۸۹
۷-۳-۲- روش بوت استرپ در آمارگیری‌های پیچیده ۹۴
۷-۳-۲-۱- نمونه‌های طبقه‌بندی شده با جایگذاری ۹۵
۷-۳-۲-۱-۱- روش ابتدایی ۹۵
۷-۳-۲-۱-۲- روش رائو و وو (Rescaling Bootstrap) ۹۶
۷-۳-۲-۲- نمونه‌گیری تصادفی ساده‌ی طبقه‌بندی شده‌ی بدون جایگذاری ۹۹
۷-۳-۲-۲-۱- روش بوت‌استرپ تطبیقی آینه‌وار ۱۰۰
۷-۳-۲-۳- نمونهٔ‌گیری بدون جایگذاری با احتمال‌های نابرابر ۱۰۲
۷-۳-۲-۳-۱- روش رائو، هارتلی- ککران ۱۰۲
۷-۳-۲-۴- نمونه‌گیری خوشه‌ای دو مرحله‌ای بدون جایگذاری ۱۰۳
۷-۳-۲-۴-۱- روش رائو و وو ۱۰۳
۷-۳-۲-۴-۲- روش سیتر ۱۰۵

۸-مقایسه‌ی روش‌ها ۱۰۷

۸-۱- معیار دقت ۱۰۷
۸-۲- معیار ملاحظات اجرایی ۱۰۹

۸-۳- معیار انعطاف‌پذیری ۱۱۰

بخش سوم: آنالیز عددی

۹- محاسبات **۱۱۵**
۹-۱- نرم‌افزارهای محاسبه‌ی برآورد واریانس ۱۱۵
۹-۲- محاسبه‌ی واریانس به‌وسیله‌ی نرم‌افزار SUDAAN ۱۱۸
۹-۲-۱- مشخص کردن طرح نمونه‌گیری برای SUDAAN ۱۱۸
۹-۲-۲- ورود داده‌ها در نرم‌افزار SUDAAN ۱۲۰
۹-۳- داده‌های شبیه‌سازی شده ۱۲۱
۹-۳-۱- برآورد واریانس به روش جک‌نایف و خطی‌سازی و BRR ۱۲۱
۹-۳-۲- مقایسه‌ی روش‌های برآورد واریانس در برآورد میانگین X و نسبتی $\frac{Y}{Z}$ ۱۳۲
۹-۴- برآورد واریانس به وسیله‌ی نرم‌افزار SUDAAN بر روی داده‌های حاصل از آمارگیری اشتغال و بیکاری فصل چهارم سال ۱۳۸۲ ۱۳۳
۹-۴-۱- برآورد واریانس به روش جک‌نایف ۱۳۳
۹-۴-۲- برآورد واریانس به روش خطی‌سازی ۱۳۵
۹-۴-۳- برآورد واریانس به روش BRR ۱۳۶

منابع **۱۳۸**

بخش اول

کلیات

فصل ۱

مفهوم خطای نمونه‌گیری

هر داده‌ی آماری، صرف‌نظر از منبع و روش جمع‌آوری، در معرض انواع مختلف خطا قرار دارد. از همین رو ارائه‌ی نتایج آمارگیری به همراه توضیحاتی در مورد کیفیت اطلاعات و محدودیت‌های آن بسیار حائز اهمیت است.

اطلاع از کیفیت داده‌ها نه تنها استفاده‌ی درست و تفسیر مناسب از داده‌ها را امکان پذیر می‌سازد بلکه به ارزیابی و اصلاح روش‌های به‌کار رفته در یک آمارگیری نیز منجر می‌شود که مورد اخیر به‌خصوص برای طرح‌های مستمر بسیار حائز اهمیت است.

هدف هر آمارگیری نمونه‌ای به‌دست آوردن برآوردهایی در مورد ویژگی‌های جامعه‌ی مورد نظر است. به‌طورکلی در فرآیند تولید برآوردها، احتمال بروز دو نوع خطا وجود دارد:

۱- خطا در اندازه‌گیری (اشکال در تعاریف، مفاهیم، طبقه‌بندی‌ها، سؤالات پرسشنامه؛ عدم تمایل یا توانایی پاسخ‌گو در ارائه‌ی اطلاعات درست؛ نامناسب بودن روش جمع‌آوری اطلاعات، کدگذاری، ادیت، پردازش داده‌ها و ...)

۲- خطا در برآورد (پوشش نادرست، روش نادرست انتخاب نمونه، بی‌پاسخی، خطای نمونه‌گیری)

هر یک از دو نوع خطای ذکر شده، شامل دو مؤلفه‌ی خطای متغیر و اریبی است. به این معنی که چنان‌چه در شرایط یکسان و با روش یکسان، نمونه‌گیری به دفعات تکرار شود، نتایج حاصل تحت تأثیر عامل‌های تصادفی و هم تحت تأثیر عامل‌های سیستماتیک می‌باشند. مؤلفه‌ی متغیر در خطا ناشی از

احتمال انتخاب نمونه‌های مختلف است که موجب بروز تفاوت بین تکرارها می‌شود. مؤلفه اریبی در خطا ناشی از شرایط حاکم بر طرح است که در تمامی تکرارها اثر کم و بیش یکسانی می‌گذارد.

زیر گروه‌های دو گروه اصلی خطا در آمارگیری‌ها را به شرح زیر می‌توان طبقه‌بندی نمود،

۱- خطاهای اندازه‌گیری

۱-۱- خطاهای مفهومی، خطا در تعاریف، مفاهیم و طبقه‌بندی‌ها و خطا در کاربرد اقلام فوق در طراحی پرسشنامه، دستورالعمل‌ها و آموزش

۱-۲- خطاهای پاسخ (اریبی پاسخ، واریانس پاسخ ساده و واریانس پاسخ همبسته)

۱-۳- خطاهای پردازش (خطاهای ادیت، خطاهای کد گذاری، خطاهای داده آمایی و خطاهای برنامه‌نویسی و ...)

۲- خطاهای برآورد

۲-۱ خطاهای پوشش (از قلم افتادگی، محدوده‌های نادرست، فهرست‌های کهنه و خطاهای انتخاب نمونه)

۲-۲ بی‌پاسخی (عدم همکاری، عدم دسترسی، غایب موقت و غیره)

۲-۳ خطاهای نمونه‌گیری (واریانس نمونه‌گیری و اریبی برآورد)

موارد ۱، ۲-۱ و ۲-۲ خطاهای غیر نمونه‌گیری نامیده می‌شوند.

تمایز بین دو مؤلفه درهر خطا یعنی خطای متغیر و اریبی از این جهت حائز اهمیت است که منابع بروز، روش‌های اندازه‌گیری، کنترل و ارزیابی این دو مؤلفه و همچنین نحوه‌ی تأثیر هر یک بر نتایج آمارگیری، متفاوت است.

اریبی‌ها می‌توانند مثبت یا منفی باشند در نتیجه اریبی‌های با علامت‌های مختلف یکدیگر را خنثی می‌کنند اما خطاهای متغیر کمیت‌های مثبتی هستند و معمولاً برای منابع مختلف جمع‌پذیرند مگر در مواقعی که بین آن‌ها همبستگی مثبت وجود داشته باشد.

اریبی‌ها را با مقایسه اطلاعات حاصل از طرح با اطلاعات موجود از منابع خارج از طرح که دارای دقت مطلوب هستند، می‌توان برآورد نمود، در صورتی که خطاهای متغیر را معمولاً با استفاده از اطلاعات طرح (با

مقایسه‌ی بین تکرارهای داخلی طرح) و در برخی موارد با مقایسه‌ی تکرارهای مستقل به‌دست آمده در شرایط یکسان، می‌توان محاسبه نمود.

در برخی از منابع بروز خطا، فقط مؤلفه‌ی اریبی (خطای پوشش)، در بعضی منابع فقط مؤلفه‌ی خطای متغیر (خطای نمونه‌گیری) و در بعضی منابع ترکیبی از هر دو مؤلفه (خطای پاسخ) وجود دارد.

بجز خطای نمونه‌گیری، خطاهای تصادفی که در مراحل جمع‌آوری و پردازش داده ها رخ می‌دهند نیز نوعی از خطاهای متغیر هستند.

با توجه به توضیحات فوق، انواع خطاهای آمارگیری را می‌توان به ۵ گروه به شرح ذیل تقسیم نمود.

۱- اریبی‌های پاسخ و پردازش
۲- واریانس پاسخ و پردازش
} خطای اندازه‌گیری

۳- خطاهای پوشش
۴- اریبی بی‌پاسخی
} خطای برآورد

۵- خطای نمونه‌گیری یا واریانس نمونه‌گیری (شامل اریبی روش برآورد)

هرچند برای کاربران داده‌ها، آگاهی از اندازه‌ی کل خطا و دو مؤلفه‌ی آن یعنی خطای متغیر و اریبی کافی است اما برای بهبود طرح نمونه‌گیری لازم است مؤلفه‌های خطا بر حسب منبع بروز مشخص شود که در عمل این کار بسیار مشکل و تقریباً ناممکن است زیرا اندازه‌گیری همه‌ی منابع خطا عملی نیست و کاملاً جمع‌پذیر هم نیستند. برای مثال خطای عدم پوشش و بی‌پاسخی معمولاً تداخل می‌یابند. واریانس‌های پاسخ و نمونه‌گیری نیز کاملاً قابل تفکیک نیستند.

به‌طورکلی چهار روش برای ارزیابی کیفیت داده‌ها مطرح است که از نقطه نظر منابع اطلاعاتی و ماهیت عملیات مورد نیاز متمایز هستند. این روش‌ها عبارتند از:

- روش‌های مبتنی بر اطلاعات حاصل از طرح
- روش‌های مبتنی بر تکرارهای مستقل داخل طرح (تکرارها در اینجا اجزایی از نمونه هستند که هر یک برآورد مستقلی از کل جامعه را ارائه می‌دهند.) از تغییرات بین این تکرارها برای اندازه‌گیری مؤلفه‌ی خطای متغیر (خطای نمونه‌گیری، خطای پاسخ) استفاده می‌شود.

- روش‌های مبتنی بر مصاحبه‌ی مجدد و سایر تکرارها یا عملیات مکمل که به عملیات اصلی طرح اضافه می‌شود. (مصاحبه‌های مجدد با روش‌های دقیق‌تر از طرح اصلی انجام شده و بر مبنای مقایسه‌ی نتایج با طرح اصلی، اریبی محاسبه می‌شود. در سایر روش‌های تکرار، طرح تحت همان شرایط تکرار می‌شود و از نتایج حاصل از تکرارهای مختلف برای اندازه‌گیری مؤلفه‌ی خطای متغیر استفاده می‌شود).
- مقایسه (در سطح خرد یا کلان) با اطلاعات منابع معتبر خارج از طرح

با توجه به اینکه در این بخش بررسی مفهوم خطاهای نمونه‌گیری مد نظر می‌باشد، در قسمت بعدی به تشریح این مطلب می‌پردازیم. در مورد مفهوم سایر انواع خطاهای آمارگیری می‌توان به مراجعی همچون ورما (۲۰۰۰) و لسلروکالسیبک (۱۹۹۲) مراجعه نمود.

۱-۱ واریانس نمونه‌گیری (خطای نمونه‌گیری)

واحدهایی که در یک نمونه‌ی خاص قرار می‌گیرند با احتمال انتخابی که در طرح نمونه‌گیری تعریف شده، انتخاب شده‌اند. به این معنی که حتی اگر اطلاعات لازم در مورد هر واحد انتخاب شده، بدون خطا مشاهده شود، نتایج حاصل از نمونه در معرض میزانی از عدم حتمیت ناشی از احتمال‌های مؤثر بر انتخاب واحدها قرار دارد. واریانس نمونه‌گیری، معیاری از این عدم حتمیت است.

توزیع برآوردهای حاصل از تمام نمونه‌های ممکن در یک طرح، توزیع نمونه‌ای برآورد نامیده می‌شود. متوسط توزیع نمونه‌ای یعنی متوسط وزنی روی تمام برآوردهای نمونه‌های ممکن که بر اساس احتمال‌های انتخاب وزن‌دهی شده‌اند، مقدار مورد انتظار نامیده می‌شود. به بیان دیگر اگر P_S احتمال و y_S برآورد حاصل از نمونه S باشند، مقدار مورد انتظار برآوردگر y عبارتست از

$$E(y)=\sum_S P_S y_S$$

است و واریانس آن برابر است با:

$$Var(y)=\sum_S P_S (y_s - E(y))^2$$

در مـواردی کـه مقـدار مـورد انـتظار روی تمـام نمونه‌های ممکن برابر مقدار واقعی Y در جامعه نیست (به دلایلـی غیر از خطاهای اندازه‌گیری) و در واقع ناشی از روش برآورد است، اریبی برآورد یا اریبی فنی[1] وجود دارد که برابر است با:

$$Bias(y) = E(y) - Y$$

ترکیب اثر واریانس و اریبی، میانگین توان دوم (MSE)[2] است که برابر است با

$$MSE(y) = \sum_S P_S (y_s - Y)^2 = Var(y) + (Bias(y))^2$$

واریانس متغیر مورد مطالعه در جامعه یعنی S^2 عبارت است

$$S^2 = \sum_j (Y_j - \bar{Y})^2 / N - 1 \quad \text{و} \quad \bar{Y} = \sum_j Y_j / N$$

جذر S^2 انحراف استاندارد یا $S.D$[3] نام دارد که، میزان تغییر پذیری عناصر جامعه از نظر صفت مورد بررسی را نشان می‌دهد. معادل این عبارت در نمونه، خطای استاندارد (SE)[4] نامیده می‌شود که میزان تغییر پذیری بیـن بـرآوردهای نمونـه‌ای صـفت مـورد بررسـی را نشان می‌دهد. به بیان دیگرSE، انحراف استاندارد توزیع نمونه‌ای و S.D. انحراف استاندارد توزیع جامعه‌ای عناصر است.

در نمونـه‌گـیری تصـادفی سـاده، SE تـنها بـه S و n (حجـم نمونـه) بسـتگی دارد. اگـر اثر پیچیدگی طرح نمونه‌گیری را با $(Deff^2)$[5] نشان دهیم.خواهیم داشت:

$$SE_{\text{طرح پیچیده}} = SE_{SRS} . Deff \qquad \text{و} \qquad SE_{SRS} = S / \sqrt{n}$$

در طرح نمونه‌گیری تصادفی ساده، برآورد میانگین و واریانس جامعه برابر است با

$$s^2 = \sum_J^n (y_i - \bar{y})^2 / n - 1 \qquad \text{و} \qquad \bar{y} = \sum_J^n y_i / n$$

---

[1] Technical Bias
[2] Mean Squared Error
[3] Standard Deviation
[4] Standard Error
[5] Design Effect

برای طرح‌های پیچیده‌ی خود وزن s^2 تقریب خوبی از برآورد واریانس جامعه می‌باشد. در طرح‌های پیچیده‌ی غیر خود وزن، برآورد میانگین و واریانس جامعه به صورت زیر می‌باشند.

$$\bar{y}_w = \frac{\sum w_i y_i}{\sum w_i} \qquad s^2 = \frac{n}{n-1} \frac{\sum w_i (y_i - \bar{Y})^2}{\sum w_i}$$

۱-۲ اهمیت آگاهی از خطای نمونه گیری

هرچند اطلاعات حاصل از هر آمارگیری در معرض منابع مختلف خطا قراردارند، آگاهی از خطای نمونه‌گیری برای تفسیر صحیح نتایج و طراحی مجدد طرح بسیار ضروری است.

گرچه خطای نمونه‌گیری فقط یک مؤلفه از کل خطای برآوردهای آمارگیری است و اغلب هم مهم‌ترین مؤلفه‌ی آن نمی‌باشد اما با توجه به این‌که آگاهی از میزان این خطا، حد پایین خطای کل را مشخص می‌کند، دانستن آن حائز اهمیت است چرا که اگر در طرحی حد پایین خطا، بزرگ باشد نتایج قابل استفاده نخواهند بود. علاوه بر این در آمارگیری‌ها معمولاً به‌غیر از برآوردهایی در سطوح کلی، برآوردهای زیر گروه‌ها نیز مورد نیاز می‌باشند و چون خطای نمونه گیری با رفتن از سطح کلی به سطح جزئی افزایش می‌یابد (بر عکس سایر خطاهای آمارگیری)، دانستن آن برای تعیین سطح مجاز ارائه‌ی اطلاعات تفصیلی و تحلیل آن‌ها لازم است.

اطلاع از خطای نمونه‌گیری همچنین برای طراحی مجدد نمونه‌گیری و ارزیابی طرح موجود ضروری است. از طرفی در یک آمارگیری عوامل متعددی مانند هزینه، چارچوب و کنترل خطاهای اندازه‌گیری مطرح می‌باشد و تصمیم‌گیری بهینه در مورد انتخاب اندازه‌ی نمونه، انتخاب نمونه، نحوه‌ی خوشه‌بندی و طبقه‌بندی، روش برآورد و... فقط با آگاهی از جزئیات تأثیر این انتخاب‌ها بر خطای نمونه‌گیری آماره‌های مورد نظر امکان پذیر می‌باشد.

با توجه به اینکه امروزه روش‌های مختلف محاسبه‌ی خطای نمونه‌گیری و نرم‌افزارهای مربوط به آن طراحی شده و به سهولت در دسترس قرار دارد، عدم ارائه‌ی اطلاعات مربوط به خطای نمونه‌گیری همراه با نتایج آمارگیری، توجیه پذیر نیست.

۱-۳ جنبه‌های عملی در انتخاب روش محاسبه‌ی خطای نمونه‌گیری

آمارگیری‌های با مقیاس وسیع (به ویژه آمارگیری خانواری)، عموماً بر مبنای یک طرح چند مرحله‌ای طبقه‌بندی شده (طرح پیچیده) انجام می‌شوند. این طرح‌ها معمولاً چند منظوره بوده و در آن انواع برآوردها از قبیل نسبت‌ها، میانگین‌ها و اختلاف بین نسبت‌ها؛ انواع واحدهای مورد بررسی از قبیل خانوار، و افراد و سطوح مختلف برآورد، برای متغیرهای متعددی موردنظر می‌باشند. بنابراین در برآورد خطای نمونه‌گیری، روشی باید مورد استفاده قرار گیرد که به جنبه‌های ذیل توجه داشته باشد:

- ساختار پیچیده طرح را در نظر بگیرد.
- به حد کافی قابل انعطاف باشد تا برای طرح‌های متفاوت بتواند به‌کار رود.
- برای کاربردهای مختلف در مقیاس وسیع و همچنین برای آماره‌ها و زیر گروه‌های متفاوت مناسب و ساده باشد.
- در برابر انحراف عملی طرح از مفروضات روش محاسباتی نظری به حد کافی استوار باشد.
- خواص آماری مطلوب از قبیل میانگین توان دوم خطای کوچک برآوردگر واریانس را داشته باشد.
- از جنبه‌ی هزینه و عملیات مورد نیاز مقرون به صرفه باشد.
- نرم‌افزار کامپیوتری مناسب برای کاربرد آن وجود داشته باشد.

فصل ۲

مفهوم آمارگیری‌های پیچیده[1]

در ادبیات آماری، تعاریف متعددی از آمارگیری پیچیده ارائه شده است که چند نمونه از آن در ذیل ارائه می‌شود.

۱- آمارگیری‌های پیچیده، به آمارگیری‌هایی اطلاق می‌شود که واریانس برآورد، تابع خطی و یا حتی تابعی معلوم از پارامترهای جامعه نیست و شامل آمارگیری‌هایی است که از طبقه‌بندی، چندین مرحله نمونه‌گیری خوشه‌ای، برآورد نسبتی یا رگرسیونی، پس طبقه‌بندی و ... استفاده می‌کنند. اغلب آمارگیری‌های مربوط به تولید آمار رسمی از این نوع هستند (له‌وی، لمی شو (۱۹۹۹)). به‌طورخلاصه می‌توان آمارگیری‌هایی را پیچیده نامیدکه روش نمونه‌گیری و یا روش برآورد آن پیچیده باشد.

۲- در یک کارگاه آموزشی که در مونترال کانادا در مورد ساختار داده‌های پیچیده در سال ۲۰۰۳ برگزار شد سرفصل‌ها عبارت بودند از:

۱) برآورد واریانس طرح‌های نمونه‌گیری پیچیده‌ی بدون جایگذاری

۲) مدل‌سازی داده‌های همبسته‌ی حاصل از آمارگیری‌های طولی

۳) مدل‌سازی چند مرحله‌ای داده‌های آمارگیری

۴) نظریه‌ی بی‌پاسخی پرسش در آمارگیری‌ها

در واقع به نظر می‌رسد در بند ۱، طرح‌های نمونه‌گیری بدون جایگذاری پیچیده تلقی می‌شوند و در بندهای بعدی مباحث جدیدتری در ارتباط با پیچیدگی بیان می‌شود.

---

[1] Complex Surveys