نشریه علوم دانشگاه تربیت معلم جلد۱۰، شماره ۱
تاریخ چاپ: بهار ۱۳۹۱

برآورد دقت برآوردگرهای غیرخطی در آمارگیری‌ها

روشنک علی‌محمدی: دانشگاه الزهرا(س)

چکیده

در این مقاله مسئلهٔ برآورد دقت برای برآوردگرهای غیرخطی بررسی می‌شود و رابطه‌ای برای محاسبهٔ واریانس کل این نوع از برآوردگرها بر حسب واریانس‌های نمونه‌گیری و غیرنمونه‌گیری تحت فرضیاتی ارائه می‌شود. همچنین با در نظر گرفتن مدل خطای پاسخ در آمارگیری‌ها و ارائهٔ برآوردگرهای مؤلفه‌های واریانس آن تحت فرضیات مدل از جمله نرمال بودن اثرهای تصادفی، روابطی برای محاسبهٔ دقت آماره‌های آمارگیری اثبات می‌شود. به‌منظور ارائهٔ کاربردی عملی برای برآورد واریانس کل آمارگیری‌ها، از دو مجموعه داده‌های نوعی استفاده شده است و دقت آمارگیری برای یک برآوردگر غیرخطی با به‌کارگیری نتایج حاصل در این مقاله، محاسبه شده است.

مقدمه

اطلاعات و نتایج حاصل از آمارگیری‌ها دارای کاربردهای متنوعی در برنامه‌ریزی‌ها هستند. بدیهی است که تصمیم‌گیری‌های موفق به اطلاعات آماری با درستی بالا نیاز دارند. لذا آگاهی کاربران از میزان درستی نتایج حاصل از آمارگیری‌ها ضروری به‌نظر می‌رسد. برای برآورد درستی لازم است دقت داده‌ها محاسبه شود. دقت رابطهٔ عکس با واریانس کل در آمارگیری‌ها دارد. برای تعیین واریانس کل آمارگیری علاوه بر برآورد واریانس نمونه‌گیری، محاسبهٔ واریانس غیرنمونه‌گیری نیز ضروری است.

در برآورد واریانس نمونه‌گیری لازم است به طرح نمونه‌گیری و شکل برآوردگر توجه شود. بخش اعظم نظریه نمونه‌گیری کلاسیک در مورد طرح‌های نمونه‌گیری ساده و برآوردگرهای خطی است، اما در عمل سایر حالت‌ها نیز به‌وفور رخ می‌دهند. بسیاری از نویسندگان از جمله کوکران (۱۹۷۷) دربارهٔ برآورد واریانس در ارتباط با برآوردگرها و طرح‌های نمونه‌گیری پایه‌ای بحث کرده‌اند [۱]. ولتر (۲۰۰۷) با در نظر گرفتن طرح‌های نمونه‌گیری پیچیده و انواع برآوردگرها به بررسی روش‌های محاسبهٔ واریانس (نمونه‌گیری) پرداخته است [۲]. در مراجع [۳] و [۴] نیز برخی از روش‌های برآورد واریانس نمونه‌گیری بررسی شده است.

گرووز (۲۰۰۴) به معرفی خطاهای غیرنمونه‌گیری در آمارگیری‌ها پرداخته است [۵]. باوداز (۲۰۰۶) به

واژه‌های کلیدی: دقت آمارگیری، واریانس کل، مدل خطای پاسخ، برآوردگر غیرخطی، تقریب تیلور.
دریافت ۸۷/۳/۳۰ پذیرش ۸۸/۱۰/۲
*نویسنده مسئول

[Downloaded from system.khu.ac.ir on 2024-11-23]

بررسی منابع خطای پاسخ بر اساس تکرار اندازه‌گیری‌ها پرداخته و توصیه‌هایی در این مورد ارائه داده است [۶]. علی‌محمدی (۱۳۸۴) ضمن بررسی انواع خطاهای آمارگیری، مدلی برای خطای پاسخ در آمارگیری‌ها پیشنهاد کرده است [۷] و در مرجع [۸] نیز به برآورد واریانس نمونه‌گیری برخی برآوردگرها در طرح‌های آماری گوناگون و بر اساس داده‌های واقعی پرداخته است.

یکی از کاربردهای مدل خطای پاسخ در برآورد واریانس غیرنمونه‌گیری است. برای محاسبهٔ واریانس غیرنمونه‌گیری می‌توان به برآورد مؤلفه‌های واریانس مدل خطای پاسخ مورد نظر پرداخت. علی‌محمدی (۱۳۸۶) به برآورد مؤلفه‌های واریانس مدل خطای پاسخ در آمارگیری‌های حضوری پرداخته است [۹].

در این مقاله، مسئلهٔ محاسبهٔ دقت برآوردگرهای غیرخطی (نسبت به مشاهدات) بررسی می‌شود و برای این نوع از برآوردگرها، رابطه‌ای برای محاسبهٔ واریانس کل آمارگیری بیان و اثبات می‌شود. علاوه بر این، محاسبهٔ واریانس کل آماره‌های غیرخطی (در دو حالت که به مقادیری از مجموعهٔ داده‌ها بستگی دارد)، بررسی شده است. به‌منظور کاربردی نمودن روابط حاصل لازم است که مؤلفه‌های واریانس مدل خطای پاسخ برآورد شوند، بدین طریق دقت آمارگیری برآوردگر مورد نظر قابل محاسبه می‌شود.

برای ارائهٔ کاربردی عملی از روابط ارائه‌شده در این مقاله، دو مجموعه از داده‌های نوعی در نظر گرفته شده است [۱۰]. در بخش کاربرد، میزان دقت آمارگیری یک برآوردگر غیرخطی از طریق بررسی واریانس نمونه‌گیری و واریانس غیرنمونه‌گیری برآوردگر با به‌کارگیری روابط حاصل، محاسبه می‌شود.

به‌دلیل فقدان و یا کهنگی چارچوب‌های فهرستی، در اکثر آمارگیری‌های نمونه‌ای از طرح‌های چندمرحله‌ای استفاده می‌شود، لذا در این مقاله، رابطه‌ها و برآوردگرها با در نظر گرفتن طرح نمونه‌گیری خوشه‌ای دو مرحله‌ای به‌دست آمده‌اند. داده‌های بررسی شده در بخش کاربرد نیز از آمارگیری دارای طرح نمونه‌گیری خوشه‌ای دو مرحله‌ای حاصل شده‌اند و برآوردگر مورد نظر نسبت به مشاهدات غیرخطی است. لذا برآورد دقت به‌روش مناسب و بر اساس طرح نمونه‌گیری و نوع برآوردگر محاسبه شده است.

واریانس کل آمارگیری

دقت به‌عنوان یک مؤلفهٔ مهم در کیفیت داده‌های آماری مورد توجه قرار دارد. واریانس کل به‌منظور تعیین میزان دقت آمارگیری محاسبه می‌شود. خطای کل در آمارگیری‌ها شامل خطاهای نمونه‌گیری و غیرنمونه‌گیری است.

[Downloaded from system.khu.ac.ir on 2024-11-23]

۱. محاسبهٔ واریانس کل برآوردگرهای غیرخطی

در این بخش روابط و برآوردگرهایی برای محاسبهٔ واریانس کل آماره‌های غیرخطی ارائه می‌شوند. برای ارائهٔ نتایج به‌صورتی‌که دارای قابلیت کاربرد در عمل باشند، مسئلهٔ برآورد واریانس کل آمارگیری در ارتباط با آمارگیری‌های عمده و پرکاربرد بررسی می‌شود.

برای توضیح نحوهٔ به‌کارگیری مدل خطای پاسخ در محاسبهٔ دقت برآوردگرهای غیرخطی، حالت زیر را در نظر می‌گیریم.

فرض کنید یك نمونهٔ خوشه‌ای دو مرحله‌ای که در هر مرحله نمونه‌ها به‌صورت تصادفی و بدون جای‌گذاری انتخاب شده‌اند، از جامعه‌ای در دسترس باشد. در مرحلهٔ اول، نمونه‌ای تصادفی به‌اندازهٔ n خوشه از جامعه‌ای شامل N خوشه استخراج می‌شود. هر خوشه شامل M_i واحد نمونه‌گیری است. در مرحلهٔ دوم نمونه‌گیری از هر خوشه m واحد به‌طور تصادفی انتخاب می‌شود. استفاده از بخشی از جامعه برای برآورد پارامتر جامعه، سبب بروز خطای نمونه‌گیری می‌شود.

با فرض اندازه‌گیری صفت مورد نظر از تمام افراد نمونه، مقدار خطای بی‌پاسخی صفر می‌شود. در صورتی‌که چارچوب نمونه‌گیری، مناسب و منطبق بر جامعهٔ هدف باشد، مقدار خطای پوشش نیز ناچیز است. همچنین با توجه به پیشرفت فناوری، امکان حذف خطای پردازش داده‌ها نیز وجود دارد. لذا خطای پاسخ تنها خطای غیرقابل اجتناب در بین خطاهای غیرنمونه‌گیری است.

برای محاسبهٔ واریانس کل در آمارگیری‌ها به مراحل تصادفیدن یعنی استخراج نمونهٔ تصادفی و مدل خطای پاسخ توجه می‌شود.

علی‌محمدی (۱۳۸۴) با بررسی نحوهٔ اجرای آمارگیری‌های حضوری مدل خطای پاسخ را بدین‌صورت ارائه کرده است:

$$y_{ijkls} = \mu_{ijkls} + A_i + B_{ij} + C_{ijk} + D_{ijkl} + R_{ijkls} \qquad (۱)$$

که در آن μ_{ijkls} مقدار واقعی، y_{ijkls} مقدار مشاهده شدهٔ sامین واحد (در استان iام مربوط به کارشناس مسئول jام و بازبین (و کدگذار) kام و پرسش‌گر lام)، A_i اثر استان iام، B_{ij} اثر کارشناس مسئول jام در استان iام، C_{ijk} اثر بازبین (و کدگذار) kام در استان iام برای کارشناس مسئول jام، D_{ijkl} اثر پرسش‌گر lام مربوط به بازبین kام و کارشناس مسئول jام در استان iام، و R_{ijkls} اثر sامین پاسخگوی منتسب به پرسش‌گر lام مربوط به بازبین kام و کارشناس مسئول jام در استان iام است.

[Downloaded from system.khu.ac.ir on 2024-11-23]

به‌دلیل این‌که در آمارگیری‌ها معمولاً از کارشناسان مسئول و بازبین‌های ثابت استفاده می‌شود، لذا اثر این عامل‌ها در مدل ثابت هستند. به‌دلیل اجرای این طرح‌ها در همهٔ استان‌ها، اثر استان نیز ثابت است و پرسش‌گرها و پاسخ‌گویان، اثرهایی تصادفی در مدل دارند.

برای بررسی مدل (۱) فرض می‌شود که اثرهای تصادفی مدل ناهم‌بسته و دارای توزیع‌های $D_{ijkl} \sim N(0,\sigma_D^2)$ و $R_{ijkls} \sim N(0,\sigma_R^2)$ هستند.

در عمل معمولاً به‌دلیل بروز خطای غیرنمونه‌گیری، مقادیر واقعی برای افراد نمونه قابل دسترس نیست. لذا از مقادیر مشاهده‌شده در نمونه به‌منظور برآورد پارامتر مورد نظر استفاده می‌شود.

فرض می‌شود که مقادیر مشاهده‌شده برای پاسخ‌گویان منتسب به پرسش‌گرهای مختلف با هم ناهم‌بسته‌اند و تنها اندازهٔ پاسخ‌گویان منتسب به یك پرسش‌گر خاص می‌توانند به‌هم وابسته باشند. هر پرسش‌گر با تعدادی از پاسخ‌گویان در خوشه‌های تخصیص‌یافته به وی مصاحبه می‌کند و تعداد پرسش‌گرها برابر L است. یعنی پاسخ‌گویان هر خوشه تنها با یک پرسش‌گر مصاحبه می‌شوند و تعداد خوشه‌های منتسب به هر پرسش‌گر برابر با h و $n=hL$ است. (با توجه به این‌که در مدل (۱)، تنها اثرهای پرسش‌گر و پاسخ‌گو تصادفی هستند، لذا این مؤلفه‌ها دارای واریانس بوده و تخصیص بیان شده کفایت می‌کند. لذا نیازی به بحث در مورد تخصیص پرسش‌گرها به سایر عمل‌گرهای آمارگیری نیست).

در این مقاله برآورد دقت آمارگیری برآوردگرهای غیرخطی مدنظر است. لذا ابتدا به بررسی این مسئله در حالت کلی پرداخته می‌شود و سپس برای یک برآوردگر غیرخطی خاص، روابط به‌صورت دقیق ارائه می‌شود.

برای نشان دادن برآوردگر غیرخطی در حالت کلی از نماد N استفاده می‌کنیم. به‌منظور محاسبهٔ واریانس کل دو مرحلهٔ تصادفیدن مطرح می‌شود. در مرحلهٔ اول تصادفیدن، نمونهٔ تصادفی طبق طرح نمونه‌گیری مورد نظر از جامعه انتخاب و به پرسش‌گرها منتسب می‌شوند و مرحلهٔ دوم تصادفی شدن ناشی از مدل خطای پاسخ است. با توجه به دو مرحلهٔ تصادفی شدن، واریانس برآوردگر به‌صورت زیر محاسبه می‌شود:

$$Var_{pz}(N) = E_p(Var_z(N \mid p)) + Var_p(E_z(N \mid p)) \qquad (۲)$$

در رابطهٔ (۲)، نماد p برای مرحلهٔ استخراج نمونهٔ تصادفی و انتساب آن به پرسش‌گرها و اندیس z برای مدل خطای پاسخ به‌کار می‌رود. لذا $E(.|p)$ و $Var(.|p)$ مقدار مورد انتظار و واریانس به‌شرط انتخاب نمونهٔ تصادفی و تخصیص آن به پرسش‌گرها را نشان می‌دهد. همچنین اندیس z در E و Var به‌ترتیب نشان‌دهندهٔ مقدار مورد انتظار و واریانس نسبت به مدل خطای پاسخ است.

[Downloaded from system.khu.ac.ir on 2024-11-23]

در صورتی که تقریب تیلور برای خطی کردن برآوردگر غیرخطی N به‌کار رود، عبارت $V_z(N \mid p)$ در جملهٔ اول رابطه (۲)، واریانس غیرنمونه‌گیری برآوردگر غیرخطی N است که در حالت کلی تابعی از مقادیر ثابت، مقادیر واقعی و مؤلفه‌های واریانس است (زیرا واریانس نسبت به مدل خطای پاسخ (z) محاسبه شده و تحت شرایط بیان شده یعنی قابل اغماض بودن سایر خطاهای غیرنمونه‌گیری می‌توان آن را با واریانس غیرنمونه‌گیری معادل دانست). لذا مقدار مورد انتظار $V_z(N \mid p)$ نسبت به مرحله p (یعنی انتخاب پاسخ‌گویان از جامعه و انتساب آن‌ها به پرسش‌گرها) ثابت است و در نتیجه جملهٔ اول رابطهٔ (۲) برابر با واریانس غیرنمونه‌گیری برآوردگر مورد نظر بر اساس مدل (۱) است.

در جملهٔ دوم رابطهٔ (۲)، با توجه به فرض صفر بودن $E(D)$ و $E(R)$ ، مقدار مورد انتظار برآوردگر N نسبت به مدل (۱) (یعنی $E_z(N \mid p)$) تابعی از مقادیر واقعی و مقادیر ثابت است، بنا بر این واریانس این عبارت برابر جملهٔ دوم رابطهٔ (۲) است که واریانس برآوردگر نسبت به مرحلهٔ p است و در نتیجه جملهٔ دوم رابطهٔ (۲) همان واریانس نمونه‌گیری برآوردگر را بیان می‌کند. لذا رابطهٔ (۲) نشان می‌دهد که واریانس کل برآوردگر غیرخطی را می‌توان با حاصل‌جمع واریانس‌های نمونه‌گیری و غیرنمونه‌گیری برآورد کرد. برای محاسبهٔ این واریانس‌ها می‌توان مقدار مورد انتظار و واریانس روابط مربوط را با استفاده از تقریب تیلور تابع برآوردگر مورد نظر محاسبه کرد.

حال برای بررسی دقیق‌تر مسئله، یک برآوردگر غیرخطی خاص را مد نظر قرار داده و نحوهٔ محاسبهٔ دقت آمارگیری برای آن ارائه می‌شود.

به‌دلیل کاربردهای فراوان برآوردگرهای نسبتی، محاسبهٔ واریانس کل این نوع از برآوردگرهای غیرخطی مورد توجه قرار گرفته است، که برای سایر آماره‌ها نیز می‌توان از روشی مشابه استفاده کرد. برآوردگر مورد نظر را که نسبت میانگین‌های نمونه‌ای دو صفت X و Y است با $T = \frac{\overline{X}}{\overline{Y}}$ نشان می‌دهیم. در این صورت رابطهٔ (۲) عبارت است از:

$$Var_{pz}(T) = E_p(Var_z(T \mid p)) + Var_p(E_z(T \mid p)) \qquad (۳)$$

با به‌کارگیری تقریب تیلور و تحت مدل (۱)، مقدار مورد انتظار و واریانس T به‌ترتیب عبارتند از:

$$E_z(T) = \frac{\overline{\mu}_{sx} + E(D) + E(R)}{\overline{\mu}_{sy} + E(D) + E(R)} = \frac{\overline{\mu}_{sx}}{\overline{\mu}_{sy}} \qquad (٤)$$

که در آن $\overline{\mu}_{sx}$ و $\overline{\mu}_{sy}$ به‌ترتیب میانگین مقادیر واقعی صفات X و Y هستند.

[Downloaded from system.khu.ac.ir on 2024-11-23]

$$
\begin{aligned}
Var_z(T=\frac{\bar{X}}{\bar{Y}}) &= \left[\frac{E_z(\bar{X})}{E_z(\bar{Y})}\right]^2\left[\frac{Var_z(\bar{X})}{E_z^2(\bar{X})}+\frac{Var_z(\bar{Y})}{E_z^2(\bar{Y})}-\frac{2Cov_z(\bar{X},\bar{Y})}{E_z(\bar{X})E_z(\bar{Y})}\right] \\
&= \left[\frac{\bar{\mu}_{sx}}{\bar{\mu}_{sy}}\right]^2\left[\frac{Var_z(\bar{X})}{\bar{\mu}_{sx}{}^2}+\frac{Var_z(\bar{Y})}{\bar{\mu}_{sy}{}^2}-\frac{2Cov_z(\bar{X},\bar{Y})}{\bar{\mu}_{sx}\bar{\mu}_{sy}}\right] \\
&= \frac{\bar{\mu}_{sx}^2}{\bar{\mu}_{sy}^2}\frac{1}{n}[\sigma_D^2(h(1-\frac{1}{m})+\frac{1}{m})+\frac{\sigma_R^2}{m}]\left[\frac{\bar{\mu}_{sx}{}^2+\bar{\mu}_{sy}{}^2-2\bar{\mu}_{sx}\bar{\mu}_{sy}}{\bar{\mu}_{sx}{}^2\bar{\mu}_{sy}{}^2}\right] \\
&= \frac{1}{n}[\sigma_D^2(h(1-\frac{1}{m})+\frac{1}{m})+\frac{\sigma_R^2}{m}](\frac{1}{\bar{\mu}_{sx}{}^2}+\frac{\bar{\mu}_{sx}{}^2}{\bar{\mu}_{sy}{}^4}-\frac{2\bar{\mu}_{sx}}{\bar{\mu}_{sy}{}^3})
\end{aligned}
\qquad (۵)
$$

زیرا

$$
\begin{aligned}
Var_z(\bar{y}) &= \frac{1}{n^2}Var_z(\sum_{a=1}^{n}(\frac{1}{m}\sum_{b=1}^{m}y_{ab})) = \frac{1}{n^2}[Var_z(\sum_{a=1}^{n}\bar{y}_{am})] \\
&= \frac{1}{n^2}(\sum_{a=1}^{n}(Var_z(\bar{y}_{am}))+\sum_{a\neq a'}\sum Cov_z(\bar{y}_{am},\bar{y}_{a'm})] \\
&= \frac{1}{n^2}[n(\sigma_D^2+\frac{\sigma_R^2}{m})+n(h-1)(\frac{m-1}{m})\sigma_D^2] \\
&= \frac{1}{n}[\sigma_D^2(h(1-\frac{1}{m})+\frac{1}{m})+\frac{\sigma_R^2}{m}]
\end{aligned}
$$

که در آن $\bar{y}$ طبق طرح خوشه‌ای دومرحله‌ای تعریف شده، n اندازهٔ خوشه‌های نمونه در مرحلهٔ اول نمونه‌گیری و m تعداد واحدهایی است که در مرحلهٔ دوم نمونه‌گیری به‌طور تصادفی از خوشه‌های مرحلهٔ اول انتخاب می‌شوند.

از رابطهٔ (۴) نتیجه می‌شود:

$$Var_p(E_z(T)) = Var_p(\frac{\bar{\mu}_{sx}}{\bar{\mu}_{sy}}) \qquad (۶)$$

رابطهٔ (۶) نشان‌دهندهٔ تغییرپذیری میانگین مقادیر واقعی نسبت به طرح نمونه‌گیری است. لذا واریانس نمونه‌گیری برآوردگر T را بیان می‌کند.

برای بررسی واریانس کل برآوردگر نسبتی ملاحظه می‌شود که جملهٔ اول عبارت سمت راست رابطهٔ (۳)، مقدار مورد انتظار رابطهٔ (۵) است که با خودش یکسان بوده و واریانس غیرنمونه‌گیری برآوردگر را نشان می‌دهد، که برای برآورد آن به‌جای $\bar{\mu}_{sx}$ و $\bar{\mu}_{sy}$ از میانگین مشاهدات در نمونه استفاده می‌شود. جملهٔ دوم سمت

[Downloaded from system.khu.ac.ir on 2024-11-23]

راست رابطهٔ (۳) واریانس نمونه‌گیری را نشان می‌دهد. لذا نتیجه می‌شود که واریانس کل برآوردگر مورد نظر تحت فرضیات بیان‌شده، برابر با مجموع واریانس نمونه‌گیری و واریانس غیرنمونه‌گیری برآوردگر است و علاوه بر این مقدار واریانس غیرنمونه‌گیری را می‌توان با استفاده از رابطهٔ (۵) به‌دست آورد.

چنان‌که از رابطهٔ (۵) ملاحظه می‌شود، به‌منظور برآورد واریانس غیرنمونه‌گیری لازم است مؤلفه‌های واریانس مربوط به اثرهای پرسش‌گر و پاسخگو برآورد شود. علی‌محمدی (۱۳۸۶) مؤلفه‌های واریانس مدل خطای پاسخ (۱) را به‌روش ماکسیمم درست‌نمایی مقید، محاسبه کرده و نشان داده است که برآوردگرهای ماکسیمم درست‌نمایی مقید مؤلفه‌های واریانس مدل (۱) (که مربوط به اثرهای تصادفی پرسش‌گر و پاسخگو هستند)، به‌ترتیب عبارتند از:

$$\hat{t} = MSR$$
$$\hat{\sigma}_D^2 = \frac{MSD - MSR}{S} \qquad (۷)$$

که در رابطهٔ (۷)، MSR و MSD به‌ترتیب میانگین توان‌های دوم اثرهای R و D در مدل (۱) و S تعداد پاسخگویان منتسب به هر پرسش‌گر است.

همچنین در صورتی‌که در رابطهٔ (۷)، $MSD\langle MSR$ باشد، برآوردگر مؤلفه‌های واریانس مدل (۱) به‌روش ماکسیمم درست‌نمایی مقید بدین صورت حاصل می‌شوند:

$$\hat{t} = \frac{SSR + SSD}{IJK(LS-1)}$$
$$\hat{\sigma}_D^2 = 0$$

با قرار دادن این مقادیر در رابطهٔ (۵)، واریانس غیرنمونه‌گیری برآوردگر T بدین‌صورت نتیجه می‌شود:

$$\hat{V}(T) = \frac{1}{n}[\hat{\sigma}_D^2(h(1-\frac{1}{m})+\frac{1}{m}) + \frac{\hat{\sigma}_R^2}{m}](\frac{1}{\bar{x}^2} + \frac{\bar{x}^2}{\bar{y}^4} - \frac{2\bar{x}}{\bar{y}^3}) \qquad (۸)$$

لذا در صورتی‌که $MSD \geq MSR$ باشد،

$$\hat{V}(T) = \frac{1}{n}[\frac{MSD-MSR}{S}(h(1-\frac{1}{m})+\frac{1}{m}) + \frac{MSR}{m}](\frac{1}{\bar{x}^2} + \frac{\bar{x}^2}{\bar{y}^4} - \frac{2\bar{x}}{\bar{y}^3}) \qquad (۹)$$

و در صورتی‌که $MSD\langle MSR$ باشد،

$$\hat{V}(T) = \frac{1}{n}[\frac{SSR+SSD}{mIJK(LS-1)}](\frac{1}{\bar{x}^2} + \frac{\bar{x}^2}{\bar{y}^4} - \frac{2\bar{x}}{\bar{y}^3}) \qquad (۱۰)$$

[Downloaded from system.khu.ac.ir on 2024-11-23]

برای محاسبهٔ واریانس غیرنمونه‌گیری برآوردگر نسبتی T می‌توان با توجه به ماهیت داده‌ها از هر یک از روابط (۹) یا (۱۰) استفاده کرد. در صورتی‌که برآورد دقت سایر آماره‌های غیرخطی مورد نظر باشد، نتایج مربوط را می‌توان با به‌کارگیری روش کلی بیان شده در این بخش، به‌دست آورد.

کاربرد

به‌منظور ارائهٔ کاربردی عملی از نتایج حاصل در بخش‌های قبل، از دو مجموعه از داده‌های نوعی مربوط به طرح هزینه و درآمد خانوار استفاده می‌شود. منبع این داده‌ها یک طرح تحقیقاتی است که مشخصات آن در مرجع [۱۰] آمده است. با توجه به نحوهٔ اجرای این طرح و هماهنگی‌های لازم در اجرای آن میزان خطای پوشش و خطای بی‌پاسخی ناچیز فرض شده است. روش نمونه‌گیری در این طرح آماری، خوشه‌ای دومرحله‌ای است. مرحلهٔ اول نمونه‌گیری مربوط به انتخاب بلوک‌ها و مرحلهٔ دوم مربوط به انتخاب خانوارهای نمونه در بلوک‌های انتخاب‌شده است. بلوک‌های نمونه با استفاده از فایل چارچوب و به‌روش سیستماتیک P. P. S. (احتمال انتخاب متناسب با تعداد خانوار) انتخاب می‌شوند. پرسش‌نامهٔ این طرح سوالات متعددی دارد. این سوالات مربوط به ریز اقلام هزینه و درآمدهای گوناگون خانوار است.

در این بررسی دقت آمارگیری یک برآوردگر غیرخطی در دو مجموعه از داده‌ها محاسبه می‌شود. برآوردگر مورد نظر نسبت هزینه‌های خوراک به درآمد کل خانوارها است.

چنان‌که در رابطG (۲) ملاحظه شد، واریانس کل برآوردگر از مجموع واریانس‌های نمونه‌گیری و غیرنمونه‌گیری قابل محاسبه است و به‌منظور تعیین واریانس غیرنمونه‌گیری می‌توان از روابط (۹) یا (۱۰) استفاده کرد.

برای برآورد واریانس نمونه‌گیری برآوردگر مورد نظر از روش جک‌نایف استفاده شده است و این مقدار بر اساس طرح نمونه‌گیری (خوشه‌ای دومرحله‌ای) برای برآوردگر غیرخطی موردنظر (یعنی نسبت هزینه‌های خوراک به درآمد کل) محاسبه می‌شود.

در مجموعه داده‌های اول، برآورد پارامتر مورد نظر برابر ۰/۶۶ و واریانس نمونه‌گیری برآوردگر ۰/۰۰۲۱ حاصل می‌شود. برای مجموعه داده‌های دوم، مقدار پارامتر ۰/۶۱ برآورد می‌شود و واریانس نمونه‌گیری آن ۰/۰۰۱ به‌دست می‌آید.

حال به بررسی واریانس غیرنمونه‌گیری برآوردگر نسبتی مورد نظر با به‌کارگیری روابط حاصل در بخش‌های قبل می‌پردازیم.

[Downloaded from system.khu.ac.ir on 2024-11-23]

برای داده‌های مجموعهٔ اول، مقدار $MSD = ۲/۲۹۴\times۱۰^{۱۵}$ و $MSR=۳/۲۸۷\times۱۰^{۱۴}$ حاصل می‌شود. لذا واریانس غیرنمونه‌گیری این برآوردگر از رابطهٔ (۹) برابر با ۰/۰۴۵۵ به‌دست می‌آید. واریانس کل آمارگیری برآوردگر مورد نظر، با توجه به مقدار واریانس نمونه‌گیری آن و با به‌کارگیری روابط ارائه‌شده، برابر با ۰/۰۴۷۶ حاصل می‌شود.

در مجموعهٔ دوم داده‌ها، مقادیر $MSD = ۱/۱۱۴\times۱۰^{۱۵}$ و $MSR =۱/۴۵۹\times۱۰^{۱۴}$ به‌دست می‌آید. لذا از رابطهٔ (۹)، مقدار واریانس غیرنمونه‌گیری برابر با ۰/۰۳۶۴ برآورد می‌شود، بنا بر این واریانس کل آمارگیری برابر با ۰/۰۳۷۴ حاصل می‌شود.

چنان‌که ملاحظه می‌شود، میزان واریانس کل آمارگیری برآوردگر غیرخطی مورد نظر برای هر دو مجموعه از داده‌ها با به‌کارگیری برآوردگرها و روابط حاصل در این مقاله، محاسبه شد. نتایج حاصل نشان می‌دهد که دقت آمارگیری در مجموعه داده‌های دوم بیش‌تر از مجموعهٔ اول در برآورد پارامتر مورد نظر است.

تعیین دقت نتایج حاصل از آمارگیری‌ها دارای اهمیت و کاربردهای فراوانی است. برای این منظور لازم است که واریانس کل آماره‌ها در آمارگیری محاسبه شود. لذا استفاده از روش ارائه‌شده و روابط حاصل در این مقاله برای برآورد دقت برآوردگرهای غیرخطی در آمارگیری‌ها پیشنهاد می‌شود.

نتیجه‌گیری

توجه به میزان کیفیت نتایج حاصل از آمارگیری‌ها به‌دلیل کاربردهای فراوان آن‌ها ضروری می‌نماید. برای محاسبهٔ کیفیت داده‌ها لازم است دقت آماره‌ها برآورد شود. دقت رابطهٔ عکس با واریانس کل آمارگیری دارد. در این مقاله نحوهٔ برآورد واریانس کل برآوردگرهای غیرخطی مورد توجه قرار گرفته است و روابطی برای محاسبهٔ دقت برآوردگرهای غیرخطی در حالت کلی ارائه شده است. علاوه بر این به‌منظور کاربردی کردن نتایج حاصل، برآوردگرها و روابط مربوط در مورد یک برآوردگر غیرخطی پرکاربرد به‌صورت دقیق‌تر تعیین شده است. برای سایر برآوردگرهای غیرخطی نیز نتایج مورد نظر با به‌کارگیری روش کلی ارائه‌شده، قابل حصول است.

به‌منظور ارائهٔ کاربردی عملی از نتایج حاصل در این مقاله، از دو مجموعه از داده‌های نوعی استفاده شده و میزان دقت یک برآوردگر غیرخطی در این دو مجموعه از داده‌ها محاسبه شده است.

تعیین دقت نتایج حاصل از آمارگیری‌ها دارای کاربردهای متعددی است. لذا به‌کارگیری نتایج و برآوردگرهای حاصل در این مقاله به‌منظور برآورد دقت آماره‌های غیرخطی در آمارگیری‌ها توصیه می‌شود.

در پایان لازم به ذکر است که این تحقیق با حمایت مالی دانشگاه الزهرا(س) انجام شده است.

[Downloaded from system.khu.ac.ir on 2024-11-23]

منابع

1. W. G. Cochran, "Sampling Techniques", 3d ed., Wiley, N. Y. (1977).

2. K. M. Wolter, "Introduction to Variance Estimation", Second Ed., Springer, N. Y. (2007).

3. P. S. Levy, "Sampling of Populations: Methods and Applications", Wiley, N. Y. (1991).

4. C. E. Sarndal, S. Lundstrom, "Estimation in Surveys with Nonresponse", Wiley, N. Y. (2005).

5. R. M. Groves, "Survey Errors and Survey Costs", Wiley, N. Y. (2004).

6. M. Bavdaz, "The Response Process in Recurring Business Surveys", Proceeding of European Conference on Quality in Survey Statistics (2006).

۷- ر. علی‌محمدی، *مدل‌سازی خطای پاسخ در آمارگیری‌های حضوری*، رساله دکترای آمار، دانشگاه تربیت مدرس، تهران (۱۳۸۴).

8- ر. علی‌محمدی، *برآورد خطای نمونه‌گیری در آمارگیری‌های پیچیده*، *هفتمین کنفرانس آمار ایران*، دانشگاه علامه طباطبایی، تهران (۱۳۸۳).

۹- ر. علی‌محمدی، *آماره‌های آزمون و برآورد مؤلفه‌های واریانس مدل خطای پاسخ در آمارگیری‌های حضوری*، مجله علوم، دانشگاه شهید چمران، اهواز (۱۳۸۶)، ش ۱۸، ص ۳۵-۴۵.

۱۰- مرکز آمار ایران، *مقایسهٔ دو روش اطلاع‌گیری از هزینه و درآمد خانوار*، گزارش طرح تحقیقاتی، مدیریت پژوهش‌های آماری، مرکز آمار ایران، تهران (۱۳۷۸).

[Downloaded from system.khu.ac.ir on 2024-11-23]