

مقایسه‌ی دقت اندازه‌گیری برآوردگر میانگین در دو طرح نمونه‌گیری متمایز

روشنک علی‌محمدی

گروه ریاضی، دانشگاه الزهراء (س)

برآورد میزان دقت نتایج حاصل از آمارگیری‌ها به دلیل کاربرد گسترده‌ی آن‌ها در زمینه‌های گوناگون از اهمیت خاصی برخوردار است. خطای اندازه‌گیری به‌عنوان خطایی غیرقابل اجتناب در آمارگیری‌ها مطرح است. لذا تعیین دقت اندازه‌گیری‌ها مورد توجه خاصی قرار دارد.

در این مقاله، مسأله‌ی تعیین دقت اندازه‌گیری برآوردگر میانگین در طرح‌های نمونه‌گیری متفاوت مورد مطالعه قرار گرفته است و رابطه‌هایی برای محاسبه‌ی دقت اندازه‌گیری برآوردگر میانگین در دو طرح نمونه‌گیری متمایز ارائه و اثبات می‌شود. همچنین با محاسبه‌ی واریانس برآوردگر برای دو طرح نمونه‌گیری مورد نظر، اثر طرح برآوردگر میانگین بررسی شده است.

واژه‌های کلیدی: اثر طرح، نمونه‌گیری خوشه‌ای چندمرحله‌ای، دقت اندازه‌گیری.

۱ مقدمه

دقت در آمارگیری‌ها رابطه‌ی عکس با واریانس کل آمارگیری دارد. خطای کل در آمارگیری‌ها به دو بخش خطای نمونه‌گیری و خطای غیرنمونه‌گیری تقسیم می‌شود. مقدار خطای نمونه‌گیری در سرشماری‌ها صفر است اما خطای غیرنمونه‌گیری در سرشماری‌ها نیز می‌تواند روی دهد.

برای محاسبه‌ی خطای نمونه‌گیری لازم است علاوه بر طرح نمونه‌گیری به فرم برآوردگر نیز توجه شود. برای مطالعه در مورد روش‌های برآورد واریانس نمونه‌گیری می‌توان به ولتر (۲۰۰۷) مراجعه کرد.

از طرفی برآورد خطای غیرنمونه‌گیری به‌سادگی امکان‌پذیر نیست و برای این منظور ابتدا باید منابع ایجاد خطا معین شود. هنسن، هورویتز و پریترکر (۱۹۶۴)، بی‌مر و لایبرگ (۲۰۰۳)، گراوز (۲۰۰۴) و سارندال و لاندستروم (۲۰۰۵) به مطالعه در مورد برخی از خطاهای غیرنمونه‌گیری پرداخته‌اند.

خطای اندازه‌گیری را می‌توان به‌عنوان اساسی‌ترین بخش از خطاهای غیرنمونه‌گیری در نظر گرفت و به‌همین دلیل، مدل‌سازی خطای اندازه‌گیری دارای کاربردهای متعددی است. باوداز (۲۰۰۶) به مطالعه در مورد منابع ایجاد خطای اندازه‌گیری در آمارگیری‌ها پرداخته است.

به‌منظور کمی‌سازی خطای اندازه‌گیری می‌توان به مدل‌بندی آن پرداخت. ماهالانویس (۱۹۴۶) برای اولین بار از مدل‌های تحلیل واریانس در مدل‌سازی خطای اندازه‌گیری استفاده کرد. در مدلی که کیش (۱۹۶۲) بررسی کرده است، واریانس کل به‌صورت مجموع واریانس پرسشگر و سایر عوامل بیان شده است. بی‌مر و تروین (۱۹۹۷) مدلی کلی برای خطای اندازه‌گیری ارائه کردند. آیهان (۲۰۰۳) مدلی را برای خطای اندازه‌گیری در آمارگیری‌های با دو مرحله‌ی مصاحبه و مصاحبه‌ی مجدد ارائه کرد.

علی‌محمدی و نواب‌پور (۱۳۸۷) مدلی برای خطای اندازه‌گیری در آمارگیری‌ها طراحی کرده است. علی‌محمدی (۱۳۸۶) با به‌کارگیری برنامه‌ریزی غیرخطی موضوع برآورد مولفه‌های واریانس مدل خطای اندازه‌گیری را مورد بررسی قرار داده است. برای مطالعه‌ی روش‌های برنامه‌ریزی غیرخطی می‌توان به لازینسکی (۲۰۰۶)، نوسدال (۲۰۰۶) و لاونبرگ (۲۰۰۸) مراجعه کرد.

به‌منظور محاسبه‌ی دقت اندازه‌گیری برآوردگرها لازم است به طرح نمونه‌گیری و برآوردگر موردنظر توجه شود. در این مقاله دقت اندازه‌گیری برآوردگر میانگین در دو طرح نمونه‌گیری تصادفی ساده و خوشه‌ای چندمرحله‌ای مورد مطالعه قرار گرفته و روابطی برای محاسبه‌ی دقت اندازه‌گیری این برآوردگر در هر دو طرح ارائه می‌شود.

اثر طرح^۱ معمولاً برای مقایسه‌ی واریانس نمونه‌گیری یک برآوردگر در دو طرح نمونه‌گیری تعریف می‌شود. در این مقاله، اثر طرح را برای واریانس اندازه‌گیری تعریف می‌کنیم. به‌عبارت دیگر اثر طرح را حاصل تقسیم واریانس اندازه‌گیری برآوردگر در طرح نمونه‌گیری مورد نظر به واریانس اندازه‌گیری آن برآوردگر در طرح نمونه‌گیری تصادفی ساده تعریف نموده‌ایم. لذا در این مقاله، اثر طرح برآوردگر میانگین در نمونه‌گیری خوشه‌ای دو مرحله‌ای نیز محاسبه شده است.

¹Design Effect

۲ برآورد واریانس اندازه‌گیری

علی‌محمدی و نواب‌پور (۱۳۸۷) مدل خطای اندازه‌گیری را با بررسی چگونگی و مراحل اجرای طرح‌های عمده‌ی آماری به صورت زیر پیشنهاد کرد:

$$y_{ijkl} = \mu_{ijkl} + A_i + B_{ij} + C_{ijk} + D_{ijkl} + R_{ijkl} \quad (1)$$

که در آن μ_{ijkl} مقدار واقعی، y_{ijkl} مقدار مشاهده‌شده‌ی s امین واحد (در استان h ام مربوط به کارشناس مسوول z ام و بازیبن (و کدگذار) k ام و پرسشگر l ام)، A_i اثر استان h ام، B_{ij} اثر کارشناس مسوول z ام در استان h ام، C_{ijk} اثر بازیبن (و کدگذار) k ام، D_{ijkl} اثر پرسشگر l ام و R_{ijkl} اثر s امین پاسخگو است. (در مورد اندیس‌ها، $j = 1, \dots, J_i$ ، $i = 1, \dots, 30$ ، $k = 1, \dots, K_{ij}$ ، $l = 1, \dots, L_{ijk}$ و $s = 1, 2, \dots, S_{ijkl}$). در این مدل، اثر استان برای سایر عوامل در مدل، اثر کارشناس مسوول برای بازیبن‌ها (و کدگذارها)، اثر بازیبن برای پرسشگرها و اثر پرسشگر برای پاسخگوها به صورت آشیانی در نظر گرفته شده است. اثرهای استان، کارشناس مسوول و بازیبن اثرهایی ثابت و پرسشگر و پاسخگو اثرهای تصادفی هستند، لذا مدل حاصل، مدل تحلیل واریانس آمیخته‌ی آشیانی است.

ابتدا تعیین دقت اندازه‌گیری در طرح نمونه‌گیری خوشه‌ای دومرحله‌ای مورد مطالعه قرار می‌گیرد.

فرض می‌شود که مقادیر مشاهده‌شده برای پاسخگویان متناسب به پرسشگرهای مختلف با هم ناهمبسته‌اند و تنها اندازه‌ی پاسخگویان متناسب به یک پرسشگر خاص می‌توانند به هم وابسته باشند و هر پرسشگر با تعدادی از پاسخگویان در خوشه‌های تخصیص‌یافته به او مصاحبه می‌کند. یعنی پاسخگویان هر خوشه تنها با یک پرسشگر مصاحبه می‌شوند و تعداد خوشه‌های متناسب به هر پرسشگر برابر با h و $n = hL$ است (به عبارت دیگر خوشه‌های نمونه‌ی s به L زیرنمونه‌ی s_1, s_2, \dots, s_L تقسیم می‌شود و هر زیرنمونه شامل h خوشه است که افراد انتخاب‌شده در آن‌ها به تصادف به یک پرسشگر متناسب می‌شوند).

با توجه به طرح نمونه‌گیری خوشه‌ای دومرحله‌ای و مدل خطای اندازه‌گیری (۱) می‌توان ثابت کرد که واریانس اندازه‌گیری برآوردگر میانگین به صورت زیر حاصل می‌شود:

$$Var_{M1}(\bar{X}) = \frac{1}{n} [\sigma_D^2 (h(1 - \frac{1}{m}) + \frac{1}{m}) + \frac{\sigma_R^2}{m}] \quad (2)$$

لذا با به کارگیری رابطه (۲) می‌توان واریانس اندازه‌گیری برآوردگر میانگین را محاسبه کرد. علی‌محمدی (۱۳۸۶) برآورد مؤلفه‌های واریانس مدل (۱) را با به کارگیری بهینه‌سازی به صورت زیر محاسبه کرد:

برای $MSD \geq MSR$

$$\hat{\sigma}_D^2 = \frac{MSD - MSR}{S}, \quad \hat{\sigma}_R^2 = SSR \quad (3)$$

و برای $MSD < MSR$

$$\hat{\sigma}_D^2 = 0, \quad \hat{\sigma}_R^2 = SSR \quad (4)$$

که در آن S تعداد پاسخگویان منتسب به هر پرسشگر، MSD مجموع توان‌های دوم اثر D یعنی $MSD = \frac{SSD}{IJK(L-1)}$ ، MSR مجموع توان‌های دوم اثر R و برابر $MSR = \frac{SSR}{IJKL(S-1)}$ است و I تعداد استان‌ها در طرح آماری مورد نظر، J تعداد کارشناسان مسوول در هر استان، K تعداد بازیبن‌های منتسب به هر کارشناس مسوول و L تعداد پرسشگرهای مربوط به هر بازیبن است.

با جایگذاری این مقادیر در رابطه (۲) می‌توان واریانس اندازه‌گیری را در طرح نمونه‌گیری خوشه‌ای چندمرحله‌ای برآورد کرد.

۳ مقایسه‌ی واریانس اندازه‌گیری در دو طرح نمونه‌گیری و اثر طرح

کیش (۱۹۶۲) اصطلاح اثر طرح را تعریف کرد. اگر فرض کنیم که یک طرح نمونه‌گیری به نمونه n واحدی از جامعه‌ای به اندازه N منجر شود، اثر طرح برای \hat{X} به صورت زیر تعریف می‌شود

$$Def f = \sigma^2 / (N^2 PQ/n)$$

که در آن $P = X/N$ و $Q = 1 - P$ است. این عبارت، واریانس نمونه‌گیری \hat{X} تحت طرح صحیح نمونه‌گیری تقسیم بر واریانس نمونه‌گیری تصادفی ساده است. در این مقاله اثر طرح را نسبت واریانس اندازه‌گیری در طرح نمونه‌گیری مورد نظر به واریانس اندازه‌گیری در طرح نمونه‌گیری تصادفی ساده تعریف می‌کنیم. لذا برای محاسبه‌ی اثر طرح به محاسبه‌ی دقت اندازه‌گیری برآوردگر میانگین در طرح نمونه‌گیری تصادفی ساده پرداخته می‌شود. در این صورت فرض می‌شود که پاسخگوها به تصادف به پرسشگرها منتسب می‌شود. همچنین پاسخگویان منتسب به هر پرسشگر با توجه به اثر آن پرسشگر به هم

وابسته‌اند و پاسخگویان برای پرسشگرهای متمایز ناهمبسته‌اند. در این صورت می‌توان نشان داد که واریانس اندازه‌گیری برآوردگر میانگین عبارت است از:

$$Var_{M\tau}(\bar{X}) = \frac{1}{nm} [S\sigma_D^2 + \sigma_R^2] \quad (5)$$

با مقایسه‌ی روابط (۲) و (۵) ملاحظه می‌شود که برای $h = 1$ ، واریانس اندازه‌گیری در دو روش نمونه‌گیری با هم برابر است. بنابراین برای $h = 1$ ، اثر طرح برابر ۱ است. اما در عمل همواره به هر پرسشگر چندین خوشه منتسب می‌شود. لذا معمولاً $h > 1$ است. در این صورت دقت اندازه‌گیری در روش خوشه‌ای دومرحله‌ای بیشتر از روش تصادفی ساده است و هر قدر h بزرگتر باشد، دقت اندازه‌گیری برآوردگر میانگین بیشتر می‌شود و در نتیجه اثر طرح کوچکتر می‌شود. لذا با به کارگیری روابط (۳) و (۴) بر حسب مجموعه داده‌ها و جایگذاری آن‌ها در رابطه (۲) و (۵) می‌توان دقت آمارگیری و اثر طرح برآوردگر میانگین را محاسبه کرد.

۴ نتیجه‌گیری

تعیین دقت نتایج حاصل از آمارگیری‌ها به دلیل کاربرد گسترده‌ی آن‌ها مورد توجه خاصی قرار دارد. خطای اندازه‌گیری بخش قابل کمی‌سازی و غیرقابل اجتناب از خطای غیرنمونه‌گیری در آمارگیری‌ها است. در این مقاله، مسأله‌ی تعیین دقت اندازه‌گیری برآوردگر میانگین در دو طرح نمونه‌گیری خوشه‌ای چندمرحله‌ای و همچنین نمونه‌گیری تصادفی ساده مورد مطالعه قرار گرفته است و رابطه‌هایی برای محاسبه‌ی دقت اندازه‌گیری برآوردگر میانگین در این دو طرح نمونه‌گیری ارائه و اثبات می‌شود. همچنین با مقایسه‌ی واریانس اندازه‌گیری برآوردگر میانگین در دو طرح نمونه‌گیری مورد نظر، اثر طرح نمونه‌گیری بر واریانس اندازه‌گیری برآوردگر میانگین محاسبه شده است.

مراجع

- [1] Ayhan, H.O. (2003). Models of Responce Error Components in Supervised Interview-Reinterview Surveys, *Journal of Applied Statistics*, 3, No. 9, 1047-1054.

- [2] Bavdaz, M. (2006). The Response Process in Recurring Business Surveys, *Proceeding of European Conference on Quality in Survey Statistics*.
- [3] Biemer, P. and Trewin, D. (1997). A Review of Measurement Error Effects on the Analysis of Survey Data, In *Survey Measurement and Process Quality*, Edited by L. Lyberg, P. Biemer, Wiley, N.Y.
- [4] Biemer, P. and Lyberg, L.E. (2003). *Introduction to Survey Quality*, Wiley, N.Y.
- [5] Groves, R.M. (2004). *Survey Errors and Survey Costs*, Wiley, N.Y.
- [6] Hansen, M., Hurwitz, W.N. and Pritzker, L. (1964). The Estimation and Interpretation of Gross Differences and The Simple Response Variance, in C. R. Rao, *Contribution of Statistics*, Calcutta: Statistical Publishing Society.
- [7] Kish, L. (1962). Studies of Interviewer Variance for Attitudinal Variables, *Journal of American Statistical Association*, **57**, 92-115.
- [8] Luenberger D.G. and Ye, Y. (2008). *Linear and Nonlinear Programming*, Third Ed., Springer.
- [9] Mahalanobis, P. (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute, *Journal of Royal Statistical Society*, **109**, 325 - 378.
- [10] Nocedal, J. and Wright, S.J. (2006). *Numerical Optimization*, Springer.
- [11] Ruszczyński, A. (2006). *Nonlinear Optimization*, USA: Princeton University Press.
- [12] Sarndal, C.E. and Lundstrom S. (2005). *Estimation in Surveys with Nonresponse*, Wiley, New York.
- [13] Wolter, K.M. (2007). *Introduction to Variance Estimation*, Springer-Verlag, New York.
- [۱۴] علی محمدی، ر. و نواب‌پور، ح.ر. (۱۳۸۷). مدل‌سازی خطای پاسخ در آمارگیری‌ها به شیوه‌ی مصاحبه‌ی حضوری، مجله پژوهشی علوم پایه دانشگاه اصفهان ۴: ۳۳، ۱۴-۱.
- [۱۵] علی محمدی، ر. (۱۳۸۶). آماره‌های آزمون و برآورد مولفه‌های واریانس مدل

خطای پاسخ در آمارگیری‌های حضوری، مجله علوم دانشگاه شهید چمران
۱۸، ۳۵-۴۵.

Archive of SID