یک برآورد جدید از نوع نمایی برای میانگین جامعهٔ متناهی در نمونه‌گیری تصادفی ساده بدون جایگذاری

شهرام یعقوب‌زاده شهرستانی [۱]

تاریخ دریافت: ۱۳۹۶/۷/۱۱
تاریخ پذیرش: ۱۳۹۷/۸/۲۴

چکیده:
در این مقاله، برای میانگین جامعهٔ متناهی در روش نمونه‌گیری تصادفی ساده بدون جایگذاری با توجه به اطلاعات کمکی، یک برآوردگر جدید از نوع نمایی ارائه می‌شود. سپس این برآوردگر جدید با چند برآوردگر دیگر میانگین جامعهٔ متناهی به وسیلهٔ دو مجموعه‌داده‌های واقعی و با استفاده از میانگین توان دوم خطاهای آنها مقایسه می‌شود.

واژه‌های کلیدی: نمونه‌گیری تصادفی ساده بدون جایگذاری، اطلاعات کمکی، برآوردگر از نوع نمایی، میانگین توان دوم خطاها.

۱ مقدمه

استفاده از اطلاعات کمکی [۲] در نظریهٔ نمونه‌گیری، به برآوردهایی از پارامترهای مجهول منجر می‌شود که از اعتمادپذیری بیشتری برخوردارند. برآورد میانگین یک جامعهٔ متناهی به‌کمک اطلاعات کمکی، تحت روش‌های برآورد نسبتی [۳]، حاصل‌ضربی [۴]، رگرسیونی و نسبتی-نمایی [۵]، مثال‌هایی از کاربرد اطلاعات کمکی در نظریهٔ نمونه‌گیری هستند. در سال‌های اخیر نویسندگانی به‌کمک اطلاعات کمکی، تحت روش‌های برآورد نسبتی و برآورد حاصل‌ضربی در روش نمونه‌گیری تصادفی ساده، میانگین جامعهٔ متناهی را برآورد کردند که به برخی از آنها مانند سیسودیا و دویودی [۵]، یوپادیا و سینگ [۱۲]، سینگ و تیلور [۶، ۸]، سینگ و دیگران [۷، ۹، ۱۰]، یاداو و کادیلار [۱۳، ۱۴] می‌توان اشاره کرد. در این مقاله در بخش دوم، انواع برآوردگرهای میانگین جامعهٔ متناهی در روش نمونه‌گیری تصادفی ساده بدون جایگذاری ارائه می‌شود. در بخش سوم، برآوردگر جدید معرفی شده و میانگین توان دوم خطای آن به دست می‌آید. در بخش چهارم با دو مجموعه‌داده‌های واقعی، برآوردگر پیشنهادی با برآوردگرهای ارائه شده در بخش دوم، مقایسه می‌شود. بخش پنجم هم به نتایج مقاله اختصاص یافته است.

۲ انواع برآوردگرهای میانگین

فرض کنید $U_1, \ldots, U_N$ واحدهای جامعهٔ متناهی به اندازهٔ N و Y بیان‌کننده متغیر تحت بررسی و X نیز بیان‌کنندهٔ متغیر کمکی باشد که دارای همبستگی قویی با صفت تحت بررسی است. فرض کنید زوج‌های $(x_1, y_1), \ldots, (x_n, y_n)$ به روش نمونه‌گیری تصادفی ساده بدون جایگذاری از آن جامعه انتخاب شوند $\bar{X}$ و $\bar{Y}$ به‌ترتیب، میانگین متغیرهای اصلی و کمکی در جامعه و $\bar{x}$ و $\bar{y}$ نیز به‌ترتیب، میانگین متغیرهای کمکی و اصلی در نمونهٔ استخراجی باشند. هرگاه همبستگی بین متغیرهای اصلی

[۱] هیأت علمی گروه آمار، دانشگاه پیام نور، تهران، ایران

[۲] Auxiliary information
[۳] Ratio estimator
[۴] Product estimator
[۵] Exponential ratio-tye estimator

و کمکی مثبت باشد، برآوردگر مناسب میانگین جامعه یعنی $\bar{Y}$، برآورد نسبتی است که به‌صورت زیر تعریف می‌شود:

$$\bar{y}_R = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) \tag{۱}$$

که با فرض $\gamma = \left(1-\frac{n}{N}\right)/n$، میانگین توان دوم خطاهای آن با توجه به سینگ و تیلور [۸] به‌صورت زیر است:

$$MSE(\bar{y}_R) = \gamma\bar{Y}^2\left(C_Y^2 - 2\rho C_X C_Y + C_X^2\right) \tag{۲}$$

به‌طوری که $C_X^2 = S_X^2/\bar{X}^2$، $C_Y^2 = S_Y^2/\bar{Y}^2$ و ρ ضریب همبستگی ساده X و Y است. همچنین S_X^2 و S_Y^2 به‌ترتیب واریانس متغیرهای کمکی و اصلی در جامعه هستند. اگر همبستگی بین متغیرهای تحت مطالعه و کمکی، منفی باشد، آن‌گاه برآورد مناسب میانگین جامعه، برآورد ضربی است که توسط رابسون [۴] معرفی شده و به‌صورت زیر است:

$$\bar{y}_P = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right) \tag{۳}$$

که میانگین توان دوم خطاهای آن به‌صورت

$$MSE(\bar{y}_P) = \gamma\bar{Y}^2\left(C_Y^2 + 2\rho C_X C_Y + C_X^2\right) \tag{۴}$$

توسط رابسون [۴] به دست آمد. بهل و توتجا [۲۸] یک برآورد از نوع نسبتی نمایی، برای میانگین جامعهٔ متناهی در روش نمونه‌گیری تصادفی ساده بدون جایگذاری به‌صورت زیر معرفی کردند:

$$\bar{y}_{BT} = \bar{y}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) \tag{۵}$$

و میانگین توان دوم خطاهای آن را به‌صورت زیر

$$MSE(\bar{y}_{BT}) = \gamma\bar{Y}^2\left(C_Y^2 - 2\rho C_X C_Y + \frac{C_X^2}{4}\right) \tag{۶}$$

به دست آمد. سینگ و دیگران [۹] و کادیلار [۲] برآوردگرهایی برای برآورد میانگین جامعه معرفی کردند که در واقع تعمیمی از رابطهٔ (۶) است و به‌ترتیب با نمادهای $\bar{y}_s$ و $\bar{y}_k$ نشان داده می‌شوند، همراه با میانگین توان دوم خطاهای آنها به‌ازای $a \in R \neq 0$ و $b \in R$ به‌صورت زیر هستند.

$$\bar{y}_S = \bar{y}\exp\left(\frac{(a\bar{X}+b)-(a\bar{x}+b)}{(a\bar{X}+b)+(a\bar{x}+b)}\right), \tag{۷}$$

$$MSE(\bar{y}_S) = \gamma\bar{Y}^2\left(C_Y^2 - 2\theta\rho C_X C_Y + \theta^2 C_X^2\right), \tag{۸}$$

$$\bar{y}_K = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right)^a \exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right), \tag{۹}$$

$$MSE(\bar{y}_K) = \gamma\bar{Y}^2\Bigg(C_Y^2 + \rho(2a+1)C_X C_Y + (a^2+a+\frac{1}{4})C_X^2\Bigg). \tag{۱۰}$$

۳ برآورد جدید میانگین

در این بخش برآوردگر جدیدی از میانگین جامعهٔ متناهی، در روش نمونه‌گیری تصادفی ساده بدون جایگذاری که در واقع تعمیمی از رابطهٔ (۶) است و در این مقاله، برآورد از نوع نمایی تعمیم‌یافته نامیده می‌شود به‌ازای $a \in R \neq 0$، به‌صورت زیر

$$\bar{y}_{GE} = \bar{y}\left(\frac{\bar{X}+\bar{x}}{2\bar{X}}\right)^a \exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) \tag{۱۱}$$

معرفی می‌شود. برای محاسبهٔ میانگین توان دوم خطاهای برآوردگر (۱۱)، $\bar{x}$ و $\bar{y}$ به‌صورت زیر در نظر گرفته می‌شوند:

$$\begin{aligned}\bar{x} &= \bar{X}(1+\varepsilon_1),\\ \bar{y} &= \bar{Y}(1+\varepsilon_0)\end{aligned} \tag{۱۲}$$

چون $\bar{y}$ یک برآوردگر نااریب $\bar{Y}$ در روش نمونه‌گیری تصادفی ساده بدون جایگذاری است، بنا بر این

$$E(\bar{y}) = \bar{Y},$$

$$Var(\bar{y}) = \gamma C_Y^2 S_X^2$$

می‌شود که با توجه به رابطهٔ (۱۲) به‌صورت زیر به دست می‌آیند:

$$\begin{aligned}E(\varepsilon_0) &= E(\varepsilon_1) = 0,\\ E(\varepsilon_0^2) &= Var(\varepsilon_0) = \gamma C_Y^2,\\ E(\varepsilon_1^2) &= Var(\varepsilon_1) = \gamma C_X^2,\\ Var(\bar{y}) &= \bar{Y}^2 Var(\varepsilon_0)\end{aligned}$$

با توجه به آن‌که $(\bar{x},\bar{y})$، میانگین‌های نمونه‌ای برای نمونهٔ تصادفی $(x_1,y_1),\dots,(x_n,y_n)$ در روش نمونه‌گیری تصادفی ساده بدون جایگذاری هستند، بنا بر این داریم:

$$Cov(\bar{x},\bar{y})=\left(\frac{1}{n}-\frac{1}{N}\right)S_{XY}.$$

در نتیجه $Cov(\varepsilon_0,\varepsilon_1)$ از رابطهٔ زیر به دست می‌آید:

$$\begin{aligned}\left(\frac{1}{n}-\frac{1}{N}\right)S_{XY}&=Cov(\bar{y},\bar{x})\\&=Cov(\bar{Y}(1+\varepsilon_0),\bar{X}(1+\varepsilon_1))\\&=\bar{X}\bar{Y}Cov(\varepsilon_0,\varepsilon_1)\end{aligned}$$

و در نتیجه $Cov(\varepsilon_0,\varepsilon_1)=\gamma\rho C_XC_Y$، که در این حالت $E(\varepsilon_0\varepsilon_1)=\gamma\rho C_XC_Y$. با استفاده از رابطهٔ (۱۲)، رابطهٔ (۱۱) به‌صورت زیر بازنویسی می‌شود:

$$\bar{y}_{GE}=\bar{Y}(1+\varepsilon_0)\left(1+\frac{\varepsilon_1}{2}\right)^a\exp\left[\frac{\varepsilon_1}{2}\left(1+\frac{\varepsilon_1}{2}\right)^{-1}\right].$$

به‌کمک بسط مک‌لورن تابع‌های $(1+\frac{\varepsilon_1}{2})^{-1}$ و نمایی، $\bar{y}_{GE}$ به‌صورت زیر به دست می‌آید:

$$\begin{aligned}\bar{y}_{GE}=&\bar{Y}(1+\varepsilon_0)\left(1+\frac{a\varepsilon_1}{2}+\frac{a(a-1)\varepsilon_1^2}{4}+\dots\right)\\&\times\exp\left[\frac{\varepsilon_1}{2}\left(1-\frac{\varepsilon_1}{2}+\frac{\varepsilon_1^2}{4}+\dots\right)\right]\end{aligned}$$

که با نادیده گرفتن توان‌های بالاتر از ۲، به‌صورت زیر بازنویسی می‌شود:

$$\begin{aligned}\bar{y}_{GE}=&\bar{Y}\left[1+\frac{(a+1)\varepsilon_1}{2}+\frac{(4a^2-2a-1)\varepsilon_1^2}{8}\right.\\&\left.+\frac{(a+1)\varepsilon_0\varepsilon_1}{2}+\varepsilon_0\right]\end{aligned}\qquad(۱۳)$$

بنا بر این با توجه به رابطهٔ (۱۳)، داریم:

$$E(\bar{y}_{GE})=\bar{Y}+\gamma\bar{Y}\left[\frac{(4a^2-2a-1)C_X^2}{8}+\frac{(a+1)\rho C_XC_Y}{2}\right]$$

با استفاده از رابطهٔ (۱۳) و با نادیده گرفتن توان‌های بالاتر از ۲، میانگین توان دوم خطاهای $\bar{y}_{GE}$ به‌صورت زیر به دست می‌آید:

$$MSE(\bar{y}_{GE})=\gamma\bar{Y}^2\left[C_Y^2+\frac{(a+1)^2}{4}C_X^2+(a+1)\rho C_XC_Y\right]$$

a به‌گونه‌ای به دست می‌آید که $MSE(\bar{y}_{GE})$ مینیمم شود. با مساوی صفر قرار دادن مشتق مرتبهٔ اول $MSE(\bar{y}_{GE})$ نسبت به a، رابطهٔ زیر به دست می‌آید:

$$a^*=-\frac{C_X+2\rho C_Y}{C_X}.$$

با قرار دادن a^* در $MSE(\bar{y}_{GE})$، مینیمم مقدار میانگین توان دوم خطاهای $\bar{y}_{GE}$ به‌صورت زیر به دست می‌آید:

$$\min_a MSE(\bar{y}_{GE})=\gamma\bar{Y}^2C_Y^2(1-\rho^2).$$

۴ مقایسه برآوردگرها

در این بخش، برآوردگر $\bar{y}_{GE}$ با برآوردگرهای $\bar{Y}$ (برآورد میانگین جامعه به روش نمونه‌گیری تصادفی ساده بدون جایگذاری)، $\bar{y}_R$، $\bar{y}_P$، $\bar{y}_{BT}$، $\bar{y}_K$ و $\bar{y}_S$ با استفاده از معیار میانگین توان دوم خطاها، مقایسه می‌شوند، تا شرط کاراتر بودن برآوردگر $\bar{y}_{GE}$ نسبت به سایر برآوردگرهای مورد مقایسه به دست آید. بنا بر این محاسبات لازم به‌صورت زیر است:

$$\begin{aligned}&MSE(\bar{y})-\min_a MSE(\bar{y}_{GE})=\gamma\bar{Y}^2\rho^2\geq 0,\qquad(۱۴)\\&MSE(\bar{y}_R)-\min_a MSE(\bar{y}_{GE})=\gamma\bar{Y}^2(C_X-\rho C_Y)^2\geq 0,\\&MSE(\bar{y}_P)-\min_a MSE(\bar{y}_{GE})=\gamma\bar{Y}^2(C_X+\rho C_Y)^2\geq 0,\\&MSE(\bar{y}_{BT})-\min_a MSE(\bar{y}_{GE})=\gamma\bar{Y}^2\left(\frac{C_X}{2}-\rho C_Y\right)^2\geq 0,\\&MSE(\bar{y}_S)-\min_a MSE(\bar{y}_{GE})=\gamma\bar{Y}^2(\theta C_X-\rho C_Y)^2\geq 0.\end{aligned}$$

با توجه به روابط به دست آمده در رابطهٔ (۱۴)، واضح است که همواره برآوردگر $\bar{y}_{GE}$ از برآوردگرهای $\bar{y}$، $\bar{y}_R$، $\bar{y}_P$، $\bar{y}_{BT}$ و $\bar{y}_S$ بهتر است. کارایی برآوردگر $\bar{y}_{GE}$ با برآوردگرهای $\bar{y}_R$، $\bar{y}_P$، $\bar{y}_{BT}$ و $\bar{y}_S$ به‌ترتیب تحت شرط‌های $\rho=C_X/C_Y$، $\rho=-C_X/C_Y$، $\rho=C_X/2C_Y$ و $\rho=\theta C_X/C_Y$ یکسان می‌شود. اما در رابطهٔ

$$\begin{aligned}MSE(\bar{y}_K)-\min_a MSE(\bar{y}_{GE})=&\gamma\bar{Y}^2\left[a^2C_X^2\right.\\&+a(C_X^2+2\rho C_XC_Y)\\&\left.+\left(\rho^2C_Y^2+\rho C_XC_Y+\frac{C_X^2}{4}\right)\right]\end{aligned}$$

معادلهٔ

$$a^2C_X^2 + a(C_X^2 + 2\rho C_X C_Y) + \left(\rho^2 C_Y^2 + \rho C_X C_Y + \frac{C_X^2}{4}\right) = 0$$

فقط دارای جواب $a = a^* = -(C_X + \rho C_Y)/2C_X$ است. بنا بر این کارایی دو برآوردگر $\bar{y}_{GE}$ و $\bar{y}_K$ یکسان می‌شود.

۵ تحلیل داده‌ها

در این بخش با استفاده از داده‌های دو جامعهٔ متناهی و با استفاده از میانگین توان دوم خطاهای برآوردگرها که در بخش سوم به دست آمدند به مقایسه برآوردگرها پرداخته می‌شود. داده‌ها در جدول ۱ آورده شده است. داده‌های جامعه اول، در شابیر و دیگران [۶] گزارش شده که در آن، تعداد درختان سیب در ۱۰۴ روستا به‌عنوان متغیر کمکی (X) و برآورد میزان تولید سیب (بر حسب تن) بر اساس انتخاب ۲۰ روستا به‌عنوان متغیر اصلی (Y) هستند. داده‌های جامعهٔ دوم، در مورتی [۳] گزارش شده که در آن متعیر کمکی (X)، تعداد کارگران و متغیر مورد مطالعه و اصلی (Y)، خروجی ۸۰ کارخانه در ناحیه‌ای به‌کمک انتخاب ۲۰ کارخانه به‌عنوان نمونه است. در هر دو جامعه، انتخاب نمونه بر اساس روش نمونه‌گیری تصادفی ساده بدون جایگذاری صورت گرفته است.

با استفاده از داده‌های جدول ۱، میانگین توان دوم خطاهای برآوردگرهای مورد مقایسه به‌ازای $a = a^*$ و $b = 2/5$، محاسبه و در جدول ۲ آورده شده است. با توجه به نتایج جدول ۲، کارایی برآوردگر پیشنهادی برای میانگین جامعهٔ متناهی، نسبت به سایر برآوردگرها بهتر و دارای کارایی یکسان با برآوردگر $\bar{y}_K$ است.

۶ نتایج

در این مقاله، یک برآورد از نوع نمایی و به‌کمک اطلاعات کمکی برای میانگین جامعهٔ متناهی، در روش نمونه‌گیری تصادفی ساده بدون جایگذاری معرفی شد و به‌کمک داده‌های واقعی و با استفاده از میانگین توان دوم خطاها، برآوردگر پیشنهادی با سایر برآوردگرهای موجود، مقایسه شد و مشخص گردید که کارایی این برآوردگر از سایر برآوردگر بهتر و با کارایی یک برآوردگر یکسان است.

جدول ۱. مقادیر آماره‌های لازم در جامعه‌ها

پارامترها	N	n	$\bar{X}$	$\bar{Y}$	ρ	C_X	C_Y
جامعهٔ اول	۱۰۴	۲۰	۱۳۹۳۱/۶۸۰	۶/۲۵۴	۰/۸۶۰	۱/۶۵۰	۱/۸۶۰
جامعهٔ دوم	۸۰	۲۰	۲/۵۸۸	۵۱/۸۲۶	۰/۹۱۵	۰/۹۴۸	۰/۳۵۴

جدول ۲. میانگین توان دوم خطاهای برآوردگرهای مورد مقایسه بر اساس داده‌های جدول ۱

برآوردگر	مقدار میانگین توان دوم خطاها در جامعهٔ اول	مقدار میانگین توان دوم خطاها در جامعهٔ دوم
$\bar{y}$	۵/۴۶۴۵۹	۲۵۲/۴۴۲۹
$\bar{y}_R$	۱/۴۲۶۹۹۲	۸۲۵/۶۹۶۴
$\bar{y}_P$	۱۸/۱۰۲۸۱	۳۲۹۹/۹۷۹
$\bar{y}_{BT}$	۲/۳۷۰۵۱۷	۱۹۴۲/۱۸۳
$\bar{y}_S$	۲/۳۷۰۶۱۵	۶۹۲/۳۱۹۴
$\bar{y}_K$	۱/۴۲۲۹۷۹	۴۱/۰۹۱۳۹
$\bar{y}_{GE}$	۱/۴۲۲۹۷۹	۴۱/۰۹۱۳۹

مراجع

[1] Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimator, *Journal of Information and Optimization Sciences*, **12**(**1**), 159-163.

[2] Kadilar, G. O. (2016). A new exponential type estimator for the population mean in simple random sampling, *Journal of Modern Applied Statistical Methods*, **15**(**2**), 207-214.

[3] Murthy, M. N. (1967). *Sampling Theory and Methods*, Calcutta, India: Statistical Publishing Society.

[4] Robson, D. S. (1957). Applications of multivariate polykays to the theory of unbiased ratio-type estimation, *Journal of American Statistical Association*, **52**(**280**), 511-522.

[5] Sisodia, B. V. S. and Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable, *Journal of the Indian Society Agricultural Statistics*, **33**(**2**), 13-18.

[6] Singh, H. P. and Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean, *Statistics in Transition*, **6**(**4**), 555-560.

[7] Singh, H. P., Tailor, R. and Kakran, M. S. (2004). An improved estimator of population mean using power transformation, *Journal of the Indian Society Agricultural Statistics*, **58**(**2**), 223-230.

[8] Singh, H. P. and Tailor, R. (2005). Estimation of finite population mean with coefficient of variation of an auxiliary character, *Statistica*, **65**(**3**), 407-413.

[9] Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2009). Improvement in estimating the population mean using exponential estimator in simple random sampling, *International Journal of Statistics and Economics*, **3**, 13-18.

[10] Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2011). Improved exponential estimator for population variance using two auxiliary variables, *Italian Journal of Pure and Applied Mathematics*, **28**, 101-109.

[11] Shabbir, J., Haq, A. and Gupta, S. (2014). A new difference-cum-exponential type estimator of finite population mean in simple random sampling, *Revista Colombiana de Estadística*, **37**(**1**), 197-209.

[12] Upadhyaya, L. N. and Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean, *Biometrical Journal*, **41**(**5**), 627-636.

[13] Yadav, S. K. and Kadilar, C. (2013a). Improved class of ratio and product estimators, *Applied Mathematics and Computation*, **219**(**22**), 10726-10731.

[14] Yadav, S. K. and Kadilar, C. (2013b). Improved exponential type ratio estimator of population variance, *Revista Colombiana de Estadística*, **36**(**1**), 145-152.