گزیده‌مطالب آماری
سال ۱۹، شماره‌ی ۱، بهار و تابستان ۱۳۸۷، صص ۴۶-۲۷

دقت براورد نسبتی در نمونه‌گیری از مجموعه‌ی رتبه‌دار

مرتضی امینی، * ناصررضا ارقامی

دانشگاه فردوسی مشهد

چکیده. نمونه‌گیری از مجموعه‌ی رتبه‌دار (Ranked set sampling) در شرایطی که اندازه‌گیری متغیر مورد نظر مشکل و یا پرهزینه است، ولی رتبه‌بندی آن در مجموعه‌های کوچک به‌سهولت انجام می‌گیرد استفاده می‌شود. در این شیوه نمونه را به‌طور تصادفی به زیرنمونه‌هایی با اندازه‌ی برابر تقسیم کرده و اعضای هر زیرنمونه به‌طور جداگانه رتبه‌بندی می‌شود. در زیرنمونه‌ی rام، آماره‌ی مرتب rامِ زیرنمونه اندازه‌گیری و ثبت می‌شود. سماوی و مطلق [۷] یک براوردگر نسبتی در نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار ارائه کرده و از آن برای براورد پارامتر نسبت استفاده کردند. کادیلار و همکاران [۴] براوردگر فوق را برای براورد میانگین جامعه با استفاده از نمونه‌گیری مجموعه‌ی رتبه‌دار به کار گرفتند. همچنین سماوی و همکاران [۸] از این براوردگر نسبتی برای براورد چندک‌ها در نمونه‌گیری مجموعه‌ی رتبه‌دار استفاده کردند. با این حال، در مقالات ذکرشده بیش‌تر به مقایسه‌ی دقت براوردگر نسبتی در شیوه‌ی نمونه‌گیری مجموعه‌ی رتبه‌دار با دقت براوردگر نسبتی در نمونه‌گیری تصادفی ساده پرداخته و به مسئله‌ی مقایسه‌ی دقت این براوردگر با میانگین نمونه و براوردگر رگرسیونی در شیوه‌ی نمونه‌گیری مجموعه‌ی رتبه‌دار کمتر توجه شده است. در این مقاله به بررسی بیش‌تر شیوه‌ی براورد میزان دقت براوردگر نسبتی در نمونه‌گیری مجموعه‌ی رتبه‌دار و همچنین مقایسه‌ی جامع این براوردگر با سایر براوردگرهای میانگین در شیوه‌ی نمونه‌گیری ذکرشده می‌پردازیم.

واژگان کلیدی: آماره‌های مرتب؛ براوردگر نسبتی؛ متغیر کمکی؛ متغیر همراه؛ نمونه‌گیری به‌روش مجموعه‌ی رتبه‌دار.

دریافت: ۱۳۸۷/۸/۱۸، پذیرش: ۱۳۸۸/۳/۱۱

* نویسنده‌ی عهده‌دار مکاتبات

۱- مقدمه

مفهوم نمونه‌گیری از مجموعه‌ی رتبه‌دار اولین بار توسط مک اینتایر [۵] معرفی شد. این شیوه در شرایطی که اندازه‌گیری متغیری خاص مشکل و یا پرهزینه است، ولی رتبه‌بندی آن در مجموعه‌های کوچک به‌سهولت انجام می‌گیرد استفاده می‌شود. در این شیوه ابتدا نمونه‌ای به اندازه‌ی $n = k^{2}$ را به‌طور تصادفی به k زیرنمونه به‌اندازه‌ی k تقسیم کرده و سپس اعضای هر زیرنمونه را به‌طور جداگانه و بر اساس اندازه‌گیری بصری یا به هر طریق ساده و کم‌هزینه‌ی دیگر که نیاز به اندازه‌گیری دقیق مقادیر نداشته باشد، رتبه‌بندی می‌کنیم. زیرنمونه‌ی rام را به‌ترتیب صعودی مرتب کرده و مقدار مرتب rام دقیقاً اندازه‌گیری و ثبت می‌شود. این فرایند تا به دست آمدن یک نمونه‌ی کامل به اندازه‌ی k ادامه پیدا می‌کند. متغیر تصادفی مرتب شده‌ی rام در زیرنمونه‌ی rام را با $X_{r:k}^{(r)}$ نشان می‌دهیم. به‌دلیل مستقل بودن زیرنمونه‌ها از یکدیگر $X_{r:k}^{(r)}$ها از هم مستقل هستند. همچنین توزیع حاشیه‌ای آن‌ها با توزیع آماره‌ی مرتب rام در یک نمونه‌ی kتایی از جامعه، $X_{r:k}$، برابر است. بدین ترتیب همان‌طور که مک اینتایر نشان داده است، میانگین نمونه‌ی به دست آمده، براوردی نااریب از میانگین جامعه خواهد بود. همچنین کارایی نسبی این براوردگر نسبت به میانگین یک نمونه‌ی تصادفی به‌اندازه‌ی k تنها کمی کوچک‌تر از $\frac{1}{2}(k+1)$ است. در واقع نمونه‌گیری از مجموعه‌ی رتبه‌دار در شرایطی که رتبه‌بندی داده‌ها آسان‌تر از اندازه‌گیری آن‌ها باشد و یا به‌صورتی بصری قابل انجام باشد بسیار مفید و کارا خواهد بود.

حال شرایطی را در نظر بگیرید که رتبه‌بندی متغیر مورد علاقه در زیرنمونه‌ها آسان نباشد، اما بتوان از یک متغیر کمکی با همبستگی زیاد با متغیر مورد علاقه که رتبه‌بندی بصری آن امکان‌پذیر است استفاده نمود. فرض کنید متغیر مورد علاقه را با Y و متغیر کمکی را با X نشان دهیم. اگر متغیرهای تصادفی مورد علاقه‌ی متناظر با $X_{r:k}^{(r)}$ها را با $Y_{[r:k]}^{(r)}$ مشخص کنیم، این متغیرها نیز از یکدیگر مستقل بوده و دارای توزیع حاشیه‌ای یکسان با متغیرهای همراه آماره‌های ترتیبی در یک نمونه‌ی kتایی از جامعه، یعنی

$Y_{[r:k]}$ها هستند. چنین شرایطی اولین بار توسط دل و کلاتر [۳] بررسی شد. آن‌ها میانگین متغیر تصادفی Y را با میانگین $Y_{[r:k]}^{(r)}$ها یعنی

$$\overline{Y}_{k:k}=k^{-1}\sum_{r=1}^{k}Y_{[r:k]}^{(r)}$$

براورد کردند. با توجه به هم‌توزیع بودن تک‌تک $Y_{[r:k]}^{(r)}$ها با $Y_{[r:k]}$ها و این‌که $Y_{[r:k]}$ها یک جایگشت از نمونه‌ی تصادفی مستقل و هم‌توزیع Y_iها با میانگین μ_Y هستند، $\overline{Y}_{k:k}$ یک براوردگر نااریب برای μ_Y است. شیوه‌ی معرفی شده توسط دل و کلاتر برای براورد واریانس [۹]، ضریب همبستگی [۱۰] و شرایطی که در آن نمونه‌گیری از X با اندازه‌های متفاوت صورت می‌پذیرد تعمیم یافته‌اند.

متغیر کمکی، X با توجه به همبستگی بالای خود با متغیر مورد علاقه، Y اطلاعات زیادی در خصوص پارامترهای جامعه‌ی Y، بخصوص μ_Y، در بر دارد. اطلاعات موجود در متغیر کمکی درباره پارامترهای توزیع متغیر مورد علاقه در صورت وجود همبستگی بالا میان متغیر کمکی و متغیر مورد علاقه، که در اکثر موارد برقرار است، بسیار مفید و قابل توجه خواهد بود. یو و لام [۱۲] به بررسی براورد رگرسیونی در نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار پرداخته‌اند.

براوردگرهای نسبتی در شرایطی که همبستگی مثبت بالایی بین متغیر کمکی و متغیر مورد علاقه وجود داشته باشد، از کارایی بالایی نسبت به براوردگر میانگین نمونه برخوردار هستند. در این مقاله به بررسی براوردگر نسبتی بر اساس مقادیر $X_{r:k}^{(r)}$ها و $Y_{[r:k]}^{(r)}$ها در نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار می‌پردازیم. بخش ۲ شامل مفاهیم اولیه است. در بخش ۳ براوردگر نسبتی در نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار برای براورد میانگین را معرفی نموده و به محاسبه‌ی اریبی و مجموع توان‌های دوم خطای تقریبی براوردگر معرفی‌شده می‌پردازیم. بخش ۴ به مقایسه‌ی براوردگر بالا با سایر براوردگرها یعنی براوردگر میانگین در نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار، براوردگر نسبتی در نمونه‌گیری تصافی ساده و براوردگر رگرسیونی در نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار اختصاص دارد. در پایان، کاربرد موارد مطرح‌شده را در توزیع نرمال دومتغیره بررسی می‌کنیم.

۲- مفاهیم اولیه

فرض کنید $\{(X_{۱},Y_{۱}),\dots,(X_{n},Y_{n})\}$ نمونه‌ی تصادفی به اندازه‌ی $n=k^{۲}$ با تابع چگالی (جرم) احتمال توأم $P_{X,Y}$ و توابع چگالی (جرم) احتمال حاشیه‌ای P_X و P_Y باشند. همچنین قرار دهید $\mu_X=E(X)$، $\sigma_X^{۲}=E(X-\mu_X)^{۲}$، $\mu_Y=E(Y)$ و $\sigma_Y^{۲}=E(Y-\mu_Y)^{۲}$. آماره‌ی مرتب rام X در یک نمونه‌ی kتایی را با $X_{r:k}$، متغیر همراه آن را با $Y_{[r:k]}$ و توابع چگالی (جرم) احتمال حاشیه‌ای آن‌ها را به‌ترتیب با f_r و $f_{[r]}$ نشان می‌دهیم. با استفاده از قضیه‌ی احتمالات کل به‌آسانی می‌توان نشان داد که

$$P_X(x)=\frac{۱}{k}\sum_{r=۱}^{k}f_r(x)\ ,\ P_Y(y)=\frac{۱}{k}\sum_{r=۱}^{k}f_{[r]}(y). \qquad (۱)$$

حال تعریف می‌کنیم

$$\mu_{(r)}=E\left(X_{r:k}\right)\ ,\ \mu_{[r]}=E\left(Y_{[r:k]}\right),$$
$$\sigma_{(r)}^{۲}=E\left(X_{r:k}-\mu_{(r)}\right)^{۲}\ ,\ \sigma_{[r]}^{۲}=E\left(Y_{[r:k]}-\mu_{[r]}\right)^{۲},$$
$$\tau_{(r)}=\mu_{(r)}-\mu_X\ ,\ \tau_{[r]}=\mu_{[r]}-\mu_Y\ ,$$
$$\bar{\alpha}_{۲(k)}=k^{-۱}\frac{\sum_{r=۱}^{k}\tau_{(r)}^{۲}}{\sigma_X^{۲}}\ ,\ \bar{\alpha}_{۲[k]}=k^{-۱}\frac{\sum_{r=۱}^{k}\tau_{[r]}^{۲}}{\sigma_Y^{۲}},$$
$$\bar{\alpha}_{p,k}=k^{-۱}\frac{\sum_{r=۱}^{k}\tau_{(r)}\tau_{[r]}}{\left(\sigma_X\sigma_Y\right)}.$$

بنا به نامساوی کوشی-شوارتز داریم $\bar{\alpha}_{p,k}^{۲}\le\bar{\alpha}_{۲(k)}\bar{\alpha}_{۲[k]}$ که در آن تساوی تحت مدل خطی (۶) برقرار است. با توجه به (۱) داریم

$$\sum_{r=۱}^{k}\mu_{(r)}=k\,\mu_X\ ,\quad \sum_{r=۱}^{k}\mu_{[r]}=k\,\mu_Y\ ,$$
$$k^{-۱}\sum_{r=۱}^{k}\sigma_{(r)}^{۲}=\sigma_X^{۲}\left(۱-\bar{\alpha}_{۲(k)}\right),\quad k^{-۱}\sum_{r=۱}^{k}\sigma_{[r]}^{۲}=\sigma_Y^{۲}\left(۱-\bar{\alpha}_{۲[k]}\right).$$

با توجه به استقلال $Y_{[r:k]}^{(r)}$ها، $\operatorname{var}\left(\bar{Y}_{k:k}\right)=k^{-۱}\sigma_Y^{۲}\left(۱-\bar{\alpha}_{۲[k]}\right)$. بنا بر این کارایی نسبی $\bar{Y}_{k:k}$ نسبت به $\bar{Y}_k$، میانگین نمونه‌ی تصادفی به اندازه‌ی k از جامعه، برابر است با

$$\mathrm{RE}=\left(۱-\bar{\alpha}_{۲[k]}\right)^{-۱}.$$

به همین ترتیب، $\bar{X}_{k:k} = k^{-1}\sum_{r=1}^{k} X_{r:k}^{(r)}$ براوردگر نااریب برای μ_X با واریانس $\operatorname{var}\left(\bar{X}_{k:k}\right) = k^{-1}\sigma_X^{2}\left(1-\bar{\alpha}_{2(k)}\right)$ است.

۳- براوردگر نسبتی

در این بخش به بررسی براوردگر نسبتی بر اساس آماره‌های مرتب و متغیرهای همراه آن‌ها در نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار می‌پردازیم. شرایطی را در نظر بگیرید که در آن μ_X، میانگین متغیر کمکی معلوم کران‌دار مثبت و دور از صفر باشد. همچنین در این مقاله فرض می‌شود که $\delta = \frac{C_x}{C_y}$ مثبت باشد، که در آن C_x و C_y به‌ترتیب ضرایب تغییر X و Y هستند. این شرط برای سادگی مقایسه‌ها در نظر گرفته شده و در نتایج کلی تغییری ایجاد نمی‌کند. علاوه بر آن، شرط یادشده، در بسیاری از موارد کاربردی در استفاده از متغیر کمکی برای به دست آوردن براوردگرهای نسبتی در نمونه‌گیری تصادفی ساده برقرار است. این براوردگر به‌صورت زیر تعریف می‌شود

$$\hat{\mu}_R - R_{k:k}\mu_X, \qquad (2)$$

که در آن $R_{k:k} = \frac{\bar{Y}_{k:k}}{\bar{X}_{k:k}}$. در ابتدا نیاز به اثبات قضیه داریم.

قضیه ۱- با استفاده از تعاریف فوق داریم

$$\operatorname{cov}\left(\bar{X}_{k:k}, \bar{Y}_{k:k}\right) = k^{-1}\operatorname{cov}\left(X, Y\right) - k^{-2}\sum_{r=1}^{k}\tau_{(r)}\tau_{[r]}.$$

برهان. با توجه به این‌که $\left(X_{r:k}^{(r)} - \mu_{(r)}\right)\left(Y_{[r:k]}^{(r)} - \mu_{[r]}\right)$ را می‌توان به‌صورت جایگشتی از $\left(X_i - \mu_{(r)}\right)\left(Y_i - \mu_{[r]}\right)$ها در نظر گرفت، لذا داریم

$$\begin{aligned} k\operatorname{cov}\left(X, Y\right) &= \sum_{r=1}^{k}\mathrm{E}\left(\left(X_r - \mu_X\right)\left(Y_r - \mu_Y\right)\right) \\ &= \sum_{r=1}^{k}\mathrm{E}\left(\left(X_r - \mu_{(r)} + \mu_{(r)} - \mu_X\right)\left(Y_r - \mu_{[r]} + \mu_{[r]} - \mu_Y\right)\right) \\ &= \sum_{r=1}^{k}\operatorname{cov}\left(X_{r:k}^{(r)}, Y_{[r:k]}^{(r)}\right) + \sum_{r=1}^{k}\tau_{(r)}\tau_{[r]} \end{aligned}$$

با توجه به مستقل بودن $X_{r:k}^{(r)}$ها از یکدیگر و به‌طور مشابه برای $Y_{[r:k]}^{(r)}$ها داریم

$$\operatorname{cov}\left(\bar{X}_{k:k},\bar{Y}_{k:k}\right)=k^{-۲}\sum_{r=۱}^{k}\operatorname{cov}\left(X_{r:k}^{(r)},Y_{[r:k]}^{(r)}\right).$$

و اثبات تمام است. ■

۱-۳- محاسبه‌ی اریبی و میانگین توان‌های دوم خطای تقریبی براوردگر نسبتی و مقایسه‌ی آن با واریانس سایر براوردگرها

براوردگر داده‌شده در (۲) همانند براوردگر نسبتی براساس نمونه‌ی تصادفی دومتغیره، اریب است. اریبی براوردگر $\hat{\mu}_R$ به‌صورت زیر قابل محاسبه است.

$$\mathrm{E}\left(R_{k:k}\right)-\frac{\mu_Y}{\mu_X}=\mathrm{E}\left(R_{k:k}\right)-\frac{\mathrm{E}\left(R_{k:k}\bar{X}_{k:k}\right)}{\mathrm{E}\left(\bar{X}_{k:k}\right)}$$
$$=\frac{-\operatorname{cov}\left(R_{k:k},\bar{X}_{k:k}\right)}{\mu_X}$$

برای به دست آوردن فرمولی تقریبی برای میزان اریبی و همچنین میانگین توان‌های دوم خطای براوردگر $\hat{\mu}_R$ فرض می‌کنیم

$$X_i=\mu_X+\varepsilon_i$$
$$Y_i=\mu_Y+\varepsilon_i',$$

که در آن ε و ε' به‌ترتیب خطاهای تصادفی X و Y حول میانگین‌های‌شان هستند. بدین ترتیب برای آماره‌های ترتیبی داریم

$$X_{r:k}=\mu_X+\varepsilon_{(r)}$$
$$Y_{[r:k]}=\mu_Y+\varepsilon_{[r]}',$$

که در آن $\varepsilon_{(r)}$ و $\varepsilon_{[r]}'$ متغیرهای ε_i و ε_i' متناظر با $X_{r:k}$ و $Y_{[:k]}$ هستند.

به‌وضوح $\bar{Y}_{k:k}=\mu_Y+\bar{\varepsilon}_{k:k}'$، $\bar{X}_{k:k}=\mu_Y+\bar{\varepsilon}_{k:k}$ و $\mathrm{E}\left(\bar{\varepsilon}_{k:k}\right)=\mathrm{E}\left(\bar{\varepsilon}_{k:k}'\right)=۰$.

همچنین به‌آسانی می‌توان نشان داد که $\operatorname{cov}\left(\varepsilon,\varepsilon'\right)=\operatorname{cov}\left(X,Y\right)$، $\mathrm{E}\left(\varepsilon_{[r]}'\right)=\tau_{[r]}$ و $\mathrm{E}\left(\varepsilon_{(r)}\right)=\tau_{(r)}$. علاوه بر آن

$$\operatorname{var}\left(\bar{\varepsilon}_{k:k}\right)=\mathrm{E}\left(\bar{X}_{k:k}-\mu_X\right)^{۲}$$
$$=k^{-۱}\sigma X^{۲}\left(۱-\bar{\alpha}_{۲(k)}\right)$$

و

$$\begin{aligned}\operatorname{cov}\left(\overline{\varepsilon}_{k:k}, \overline{\varepsilon}'_{k:k}\right) &= k^{-۲}\sum_{r=۱}^{k}\operatorname{cov}\left(\varepsilon_{(r)}, \varepsilon'_{[r]}\right)\\ &= k^{-۱}\operatorname{cov}\left(\varepsilon, \varepsilon'\right) - k^{-۲}\sum_{r=۱}^{k}\mathrm{E}\left(\varepsilon_{(r)}\right)\mathrm{E}\left(\varepsilon'_{[r]}\right)\\ &= k^{-۱}\operatorname{cov}\left(X, Y\right) - k^{-۲}\sum_{r=۱}^{k}\tau_{(r)}\tau_{[r]}.\end{aligned}$$

بنا بر این

$$\begin{aligned}R_{k:k} &= \frac{\mu_Y + \overline{\varepsilon}'_{k:k}}{\mu_X + \overline{\varepsilon}_{k:k}}\\ &= \frac{\mu_Y}{\mu_X}\left(۱+\frac{\overline{\varepsilon}'_{k:k}}{\mu_Y}\right)\left(۱+\frac{\overline{\varepsilon}_{k:k}}{\mu_X}\right)^{-۱}.\end{aligned}$$

با توجه به خنثی شدن جملات مثبت و منفی خطا در $\overline{\varepsilon}_{k:k}$ و همچنین کران‌دار و دور از صفر بودن μ_X می‌توان فرض کرد که $\frac{\overline{\varepsilon}_{k:k}}{\mu_X} < ۱$. بنا بر این با در نظر گرفتن جملات ابتدائی بسط $\left(۱+\frac{\overline{\varepsilon}_{k:k}}{\mu_X}\right)^{-۱}$ می‌توان نوشت

$$\begin{aligned}R_{k:k} &= \frac{\mu_Y}{\mu_X}\left(۱+\frac{\overline{\varepsilon}'_{k:k}}{\mu_Y}\right)\sum_{i=۰}^{\infty}(-۱)^i\left(\frac{\overline{\varepsilon}_{k:k}}{\mu_X}\right)^i\\ &\approx \frac{\mu_Y}{\mu_X}\left(۱+\frac{\overline{\varepsilon}'_{k:k}}{\mu_Y}-\frac{\overline{\varepsilon}_{k:k}}{\mu_X}+\left(\frac{\overline{\varepsilon}_{k:k}}{\mu_X}\right)^۲-\frac{\overline{\varepsilon}_{k:k}\overline{\varepsilon}'_{k:k}}{\mu_X\mu_Y}\right).\end{aligned}$$

بدین ترتیب اریبی تقریبی $R_{k:k}$ به‌صورت زیر به دست می‌آید

$$\begin{aligned}
\mathrm{B}\left(R_{k:k}\right) &= \mathrm{E}\left(R_{k:k}\right)-\frac{\mu_Y}{\mu_X} \\
&\approx \frac{\mu_Y}{\mu_X}\left(\mathrm{E}\left(\frac{\bar{\varepsilon}_{k:k}}{\mu_X}\right)^{۲}-\frac{۱}{\mu_X\,\mu_Y}\operatorname{cov}\left(\bar{\varepsilon}_{k:k},\bar{\varepsilon}'_{k:k}\right)\right) \\
&= \frac{\mu_Y}{\mu_X}\Big[\frac{۱}{\mu_X{}^{۲}}\left(k^{-۱}\sigma_X{}^{۲}-k^{-۲}\sum_{r=۱}^{k}\tau_{(r)}{}^{۲}\right) \\
&\qquad -\frac{۱}{\mu_X\,\mu_Y}\left(k^{-۱}\rho\sigma_X\,\sigma_Y-k^{-۲}\sum_{r=۱}^{k}\tau_{(r)}\tau_{[r]}\right)\Big] \\
&= \frac{۱}{k}\frac{\mu_Y}{\mu_X}\left(C_X^{۲}\left(۱-\bar{\alpha}_{۲(k)}\right)-C_X C_Y\left(\rho-\bar{\alpha}_{p,k}\right)\right).
\end{aligned}$$

همچنین میانگین توان‌های دوم خطای $R_{k:k}$ برابر است با

$$\begin{aligned}
\mathrm{MSE}\left(R_{k:k}\right) &\approx \left(\frac{\mu_Y}{\mu_X}\right)^{۲}\mathrm{E}\left(\frac{\bar{\varepsilon}'_{k:k}}{\mu_Y}-\frac{\bar{\varepsilon}_{k:k}}{\mu_X}\right)^{۲} \\
&= \left(\frac{\mu_Y}{\mu_X}\right)^{۲}\left(\frac{۱}{\mu_Y{}^{۲}}\left(k^{-۱}\sigma_Y{}^{۲}-k^{-۲}\sum_{r=۱}^{k}\tau_{[r]}{}^{۲}\right)+\frac{۱}{\mu_X{}^{۲}}\left(k^{-۱}\sigma_X{}^{۲}-k^{-۲}\sum_{r=۱}^{k}\tau_{(r)}{}^{۲}\right)\right. \\
&\qquad \left.-\frac{۲}{\mu_X\,\mu_Y}\left(k^{-۱}\rho\sigma_X\,\sigma_Y-k^{-۲}\sum_{r=۱}^{k}\tau_{(r)}\tau_{[r]}\right)\right)
\end{aligned}$$

در نتیجه

$$\mathrm{MSE}\left(\hat{\mu}_R\right)\approx\frac{\sigma_Y{}^{۲}}{k}\left(۱-\bar{\alpha}_{۲[k]}+\delta^{۲}\left(۱-\bar{\alpha}_{۲(k)}\right)-۲\delta\left(\rho-\bar{\alpha}_{p,k}\right)\right). \qquad (۳)$$

۴- مقایسه

در این بخش بـه مقایـسه‌ی براوردگـر نـسبتی (۲) بـا برخـی دیگـر از براوردگرهـای μ_Y می‌پردازیم.

۴-۱- مقایسه با براوردگر نسبتی در نمونه‌گیری تصادفی ساده و براوردگر میانگین در نمونه‌گیری از مجموعه‌ی رتبه‌دار

با استفاده از (۳)، $\hat{\mu}_R$ کاراتر از براوردگر نسبتی در نمونه‌گیری تصادفی ساده است اگر

$$\delta^{۲}\bar{\alpha}_{۲(k)} + \bar{\alpha}_{۲[k]} - ۲\delta\bar{\alpha}_{p,k} > ۰. \qquad (۴)$$

از آن‌جا که $\bar{\alpha}_{۲(k)} > ۰$ و $\bar{\alpha}_{p,k}{}^{۲} \le \bar{\alpha}_{۲(k)}\bar{\alpha}_{۲[k]}$، شرط (۴) برقرار است مگر $\bar{\alpha}_{p,k}{}^{۲} = \bar{\alpha}_{۲(k)}\bar{\alpha}_{۲[k]}$. همان‌طور که در زیربخش بعدی خواهیم دید چنین شرایطی تحت مدل خطی بدون عرض از مبدأ برقرار هستند.

همچنین $\hat{\mu}_R$ کاراتر از $\bar{Y}_{k:k}$ است اگر

$$\delta^{۲}\left(۱-\bar{\alpha}_{۲(k)}\right) - ۲\delta\left(\rho - \bar{\alpha}_{p,k}\right) < ۰,$$

یا

$$\rho > \frac{\delta}{۲}\left(۱-\bar{\alpha}_{۲(k)}\right) + \bar{\alpha}_{p,k}. \qquad (۵)$$

کارایی نسبی $\hat{\mu}_R$ نسبت به $\bar{Y}_{k:k}$ تقریباً برابر است با

$$\mathrm{RE} \approx \left(۱ + (\delta)^{۲}\frac{۱-\bar{\alpha}_{۲(k)}}{۱-\bar{\alpha}_{۲[k]}} - ۲(\delta)\frac{\rho-\bar{\alpha}_{p,k}}{۱-\bar{\alpha}_{۲[k]}}\right)^{-۱}.$$

۴-۲- یک مدل خطی

در این زیربخش به بررسی خصوصیات $\hat{\mu}_R$ تحت مدل خطی

$$Y_i = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}\left(X_i - \mu_X\right) + \varepsilon_i \qquad (۶)$$

می‌پردازیم. به‌آسانی می‌توان نشان داد که تحت (۶)

$$\frac{\tau_{[r]}}{\sigma_Y} = \rho\frac{\tau_{(r)}}{\sigma_X}.$$

بنا بر این

$$\bar{\alpha}_{p,k} = \rho\bar{\alpha}_{2(k)},$$
$$\bar{\alpha}_{2[k]} = \rho^{2}\bar{\alpha}_{2(k)}.$$

لذا در این حالت میانگین توان‌های دوم خطای $\hat{\mu}_R$ تقریباً برابر است با

$$\mathrm{MSE}\left(\hat{\mu}_R\right) \approx \frac{\sigma_Y{}^{2}}{k}\left(\left(1-\rho^{2}\bar{\alpha}_{2(k)}\right)+\delta^{2}\left(1-\bar{\alpha}_{2(k)}\right)-2\rho\delta\left(1-\bar{\alpha}_{2(k)}\right)\right). \qquad (۷)$$

همچنین

$$\mathrm{B}\left(R_{k:k}\right)=\frac{1}{k}\frac{\mu_Y}{\mu_X}\left(1-\bar{\alpha}_{2(k)}\right)\left(C_X^{2}-\rho C_X C_Y\right). \qquad (۸)$$

۱-۲-۴- مقایسه با براوردگر رگرسیونی

یو و لام [۱۲] خصوصیات براوردگر رگرسیونی به‌شکل

$$\hat{Y}_{reg} = \bar{Y}_{k:k} + \hat{\beta}\left(\mu_X - \bar{X}_{k:k}\right),$$

که در آن

$$\hat{\beta} = \sum_{r=1}^{k}\left(X_{r:k}^{(r)} - \bar{X}_{k:k}\right)\left(Y_{[r:k]}^{(r)} - \bar{Y}_{k:k}\right)$$

را تحت مدل خطی (۶) مطالعه نمودند. آن‌ها نشان دادنـد کـه براوردگـر فـوق بـرای μ_Y نااریب بوده و واریانس آن برابر است با

$$\mathrm{var}\left(\hat{Y}_{reg}\right) = \frac{\sigma_Y^{2}}{k}\left(1-\rho^{2}\right)\left[1+\eta_z\right], \qquad (۹)$$

که در آن

$$\eta_z = \mathrm{E}\left(\frac{\bar{Z}_{k:k}^{2}}{S_{Z_k}^{2}}\right),$$

که

$$\bar{Z}_{K:k} = k^{-1}\sum_{r=1}^{k}\left(\frac{X_{r:k}^{(r)}-\mu_X}{\sigma_X}\right) = k^{-1}\sum_{r=1}^{k} Z_{r:k}^{(r)}$$

و

$$S_{Z_k}^{2} = k^{-1}\sum_{r=1}^{k}\left(Z_{r:k}^{(r)} - \bar{Z}_{k:k}\right)^{2}.$$

در نتیجه بنا بر (۷) و (۹)، $\hat{\mu}_R$ کاراتر از $\hat{Y}_{reg}$ است اگر

$$\Omega_{\rho^{۲}} + \Lambda_{\rho} + \Gamma < ۰, \qquad (۱۰)$$

که در آن

$$\Omega = ۱ - \bar{\alpha}_{۲(k)} + \eta_z ,$$
$$\Lambda = -۲(\delta)\left(۱ - \bar{\alpha}_{۲(k)}\right)$$

و

$$\Gamma = (\delta)^{۲}\left(۱ - \bar{\alpha}_{۲(k)}\right) - \eta_z .$$

نامساوی (۱۰) در زیربخش ۴-۲-۴ بررسی خواهد شد. کارایی نسبی $\hat{\mu}_R$ به $\hat{Y}_{reg}$ تقریباً برابر است با

$$\mathrm{RE} \approx \left(۱ - \rho^{۲}\right)\left(۱ + \eta_z\right)\left(\left(۱ - \rho^{۲}\bar{\alpha}_{۲(k)}\right) + \delta^{۲}\left(۱ - \bar{\alpha}_{۲(k)}\right) - ۲\rho\delta\left(۱ - \bar{\alpha}_{۲(k)}\right)\right)^{-۱}. \qquad (۱۱)$$

۴-۲-۲- مقایسه با سایر براوردگرها

از (۷) نتیجه می‌شود که $\hat{\mu}_R$ کاراتر از براوردگر نسبتی در نمونه‌گیری تصادفی ساده است اگر

$$\delta^{۲} - ۲\rho\delta + \rho^{۲} > ۰. \qquad (۱۲)$$

نامساوی اخیر همواره برقرار است مگر $\delta = \rho$. همچنین تحت مدل خطی، نامساوی (۵) به‌صورت

$$\rho > \frac{\delta}{۲}\left(۱ - \bar{\alpha}_{۲(k)}\right) + \rho\bar{\alpha}_{۲(k)}$$

در می‌آید که معادل است با

$$\rho > \frac{\delta}{۲}. \qquad (۱۳)$$

شرط (۱۳) همان شرط ترجیح براوردگر نسبتی به براوردگر میانگین در نمونه‌گیری تصادفی ساده است. علاوه بر این داریم

$$\mathrm{RE} \approx \left(1 + (\delta)^2 \frac{1-\bar{\alpha}_{2(k)}}{1-\rho^2 \bar{\alpha}_{2[k]}} - 2\rho(\delta) \frac{1-\bar{\alpha}_{2(k)}}{1-\rho^2 \bar{\alpha}_{2(k)}} \right)^{-1}. \qquad (۱۴)$$

۳-۲-۴- مدل خطی بدون عرض از مبدأ

تحت مدل خطی (۶) بدون عرض از مبدأ داریم $\rho = \delta$. در نتیجه $\mathrm{B}(R_{k:k}) \approx 0$ و

$$\mathrm{MSE}(\hat{\mu}_R) \approx \frac{\sigma_Y^{\ 2}}{k}(1-\rho^2),$$

که با توجه به این‌که $\eta_z > 0$، از $\mathrm{var}(\hat{Y}_{reg})$ در (۹) کوچک‌تر است. شرط (۱۳)، در این حالت به‌وضوح همواره برقرار است.

از طرف دیگر، تحت مدل خطی بدون عرض از مبدأ نامساوی (۱۲) به تساوی تبدیل می‌شود و در نتیجه کارایی براوردگر نسبتی در نمونه‌گیری از مجموعه‌ی رتبه‌دار و نمونه‌گیری تصادفی ساده با هم برابرند.

از بحث اخیر می‌توان چنین نتیجه گرفت که اگرچه تحت مدل خطی بدون عرض از مبدأ براوردگر نسبتی تقریباً نااریب و کاراتر از براوردگر رگرسیونی و میانگین است، در این حالت، نیازی به استفاده از روش نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار نیست و بهتر است که میانگین را با استفاده از براوردگر نسبتی در نمونه‌گیری تصادفی ساده براورد نمود. بنا بر این بایستی به مطالعه‌ی شرایطی که براوردگر نسبتی از هردو براوردگر رگرسیونی و میانگین در یک مدل خطی با عرض از مبدأ کاراتر است بپردازیم. چنین شرایطی در زیربخش بعد مورد بررسی قرار گرفته و بعداً در مثال توزیع نرمال دومتغیره به بررسی دقیق‌تر آن پرداخته شده است.

۴-۲-۴- بحث

در این زیربخش مدل خطی (۶) با عرض از مبدأ را در نظر می‌گیریم، یعنی $\rho \neq \delta$. نامساوی (۱۰) را در نظر گرفته و قرار دهید

$$\Delta' = \Lambda^2/4 - \Omega\Gamma = \eta_z \left[\left(1-\bar{\alpha}_{2(k)}\right)\left(1-(\delta^2)\right) + \eta_z \right].$$

جداول آماری $\mu_{(r)}$ نشان می‌دهند که تقریباً برای تمامی توزیع‌های استاندارد برای X، مثل نرمال، نمایی، یکنواخت و غیره، $\bar{\alpha}_{۲(k)} <۱$ ([۱] را ببینید). همچنین $\eta_z >۰$ و در نتیجه $\Omega >۰$. حال سه حالت زیر را در نظر می‌گیریم.

۱- $\delta = ۱$: در این حالت $\Delta' = \eta_z^۲ > ۰$. بنا بر این $\hat{\mu}_R$ از $\hat{Y}_{reg}$ کاراتر است اگر

$$\rho > \frac{۱-\bar{\alpha}_{۲(k)}-\eta_z}{۱-\bar{\alpha}_{۲(k)}+\eta_z}.$$

۲- $\delta <۱$: داریم $\Delta' = \eta_z\left[\left(۱-\bar{\alpha}_{۲(k)}\right)\left(۱-\delta^۲\right)+\eta_z\right] >۰$. بنا بر این $\hat{\mu}_R$ به‌ازای $\rho_{-۱} < \rho < \rho_۱$، که در آن

$$\rho_i = \frac{\delta\left(۱-\bar{\alpha}_{۲(k)}\right)+i\sqrt{\Delta'}}{۱-\bar{\alpha}_{۲(k)}+\eta_z}, \qquad i=-۱,۱$$

از $\hat{Y}_{reg}$ کاراتر است.

۳- $\delta >۱$: اگر $۱<\delta<\sqrt{۱+\frac{\eta_z}{۱-\bar{\alpha}_{۲(k)}}}$، داریم $\Delta' >۰$ و بنا بر این $\hat{\mu}_R$ به‌ازای $\rho_{-۱} < \rho < \rho_۱$ از $\hat{Y}_{reg}$ کاراتر است. اگر $\delta = \sqrt{۱+\frac{\eta_z}{۱-\bar{\alpha}_{۲(k)}}}$، داریم $\Delta' = ۰$، در این صورت $\hat{\mu}_R$ از $\hat{Y}_{reg}$ کاراتر است، مگر برای $\rho = \delta^{-۱}$. نهایتاً اگر $\delta > \sqrt{۱+\frac{\eta_z}{۱-\bar{\alpha}_{۲(k)}}}$، داریم $\Delta' <۰$ و $\hat{Y}_{reg}$ همواره از $\hat{\mu}_R$ کاراتر است.

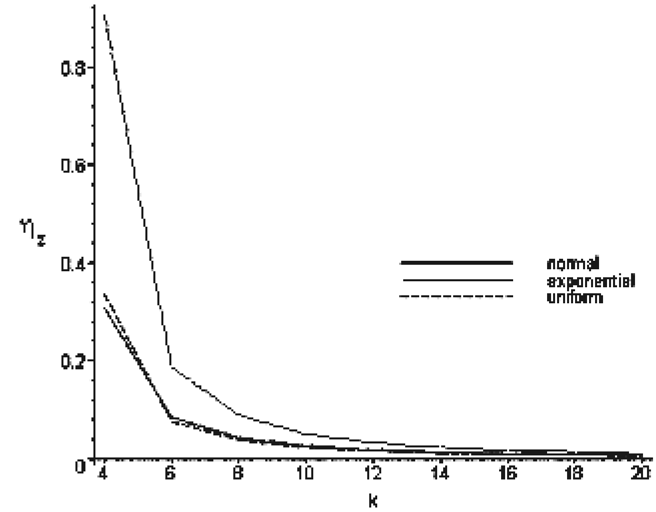
می‌توان به‌آسانی نشان داد که به‌ازای تمام مقادیر $\delta < ۱+\frac{\eta_z}{۱-\bar{\alpha}_{۲(k)}}$، $۱> \rho_۱$. همچنین، به‌ازای $۱-\frac{\eta_z}{۱-\bar{\alpha}_{۲(k)}} < \delta < ۱$ داریم $\rho_{-۱} > \frac{۱-\bar{\alpha}_{۲(k)}-\eta_z}{۱-\bar{\alpha}_{۲(k)}+\eta_z}$. به‌ازای $۰<\delta<۱-\frac{\eta_z}{۱-\bar{\alpha}_{۲(k)}}$ داریم $\rho_{-۱} < \frac{۱-\bar{\alpha}_{۲(k)}-\eta_z}{۱-\bar{\alpha}_{۲(k)}+\eta_z}$.

از بحث بالا می‌توان نتیجه گرفت که برای $\delta = ۱$، $\hat{\mu}_R$ به‌ازای بازه‌ی بزرگ‌تری از مقادیر ρ از $\hat{Y}_{reg}$ کاراتر است. طول این بازه با افزایش δ از ۱ به $\sqrt{۱+\frac{\eta_z}{۱-\bar{\alpha}_{۲(k)}}}$ و یا کاهش آن از ۱ به $۱-\frac{\eta_z}{۱-\bar{\alpha}_{۲(k)}}$ کاهش می‌یابد.

خارج از این بازه‌ی مقادیر δ، رفتار براوردگرها تغییر می‌کند. به‌ازای مقادیر $\delta \geq \sqrt{۱+\frac{\eta_z}{۱-\bar{\alpha}_{۲(k)}}}$ طول بازه‌ی مقادیری از ρ که به‌ازای آن‌ها $\hat{\mu}_R$ از $\hat{Y}_{reg}$ کاراتر است، به

صفر می‌رسد و در نتیجه براوردگر رگرسیونی ترجیح داده می‌شود. اما به‌ازای مقادیر $0 < \delta < 1 - \frac{\eta_z}{1-\bar{\alpha}_{2(k)}}$ بازه‌ی مقادیر ρ که به‌ازای آن‌ها $\hat{\mu}_R$ از $\hat{Y}_{reg}$ کاراتر است، به سمت چپ انتقال یافته و طول آن افزایش می‌یابد. این به این معنی است که به‌ازای این مقادیر δ، براوردگر نسبتی حتی برای مقادیر همبستگی مثبت کوچک‌تر نیز ترجیح داده می‌شود. با میل کردن δ به صفر بازه‌ی مقادیر ρ که به‌ازای آن‌ها $\hat{\mu}_R$ از $\hat{Y}_{reg}$ کاراتر است، به بازه‌ی متقارن $\left(-\sqrt{\frac{\eta_z}{1-\bar{\alpha}_{2(k)}+\eta_z}}, \sqrt{\frac{\eta_z}{1-\bar{\alpha}_{2(k)}+\eta_z}}\right)$ نزدیک می‌شود. در این‌جا باید یادآور شویم که نتایج بیان‌شده، بر اساس میانگین توان‌های دوم خطای تقریبی $\hat{\mu}_R$ می‌باشند.

برای مقادیر بزرگ k، جداول آماری نشان می‌دهد که $\bar{\alpha}_{2(k)}$ به ۱ میل می‌کند. همچنین یک شبیه‌سازی رایانه‌ای با ۱۰۰۰۰۰ تکرار (شکل۱) نشان می‌دهدکه η_z به‌سرعت به صفر میل می‌کند. بنا بر این وقتی $k \to \infty$، Ω بسیار آهسته به صفر میل می‌کند نمی‌توانیم برای مقادیر بزرگ k، (۱۰) را در حالت خطی ان در نظر بگیریم.



شکل ۱- منحنی η_z در مقابل k در سه توزیع متداول

برای حالت $\delta = 1$، $\frac{1-\bar{\alpha}_{2(k)}-\eta_z}{1-\bar{\alpha}_{2(k)}+\eta_z}$ به ۱ میل می‌کند و در نتیجه برای مقادیر بزرگ k، براوردگر رگرسیونی ترجیح داده می‌شود. با میل کردن η_z به صفر، حدود ρ_{-1} و ρ_1 بر

هم منطبق می‌شوند. گرچه از این بحث می‌توان نتیجه گرفت که برای مقادیر بزرگ k، براوردگر رگرسیونی کاراتر از براوردگر نسبتی است اما از آن‌جا که هر دو براوردگر سازگار هستند به‌ازای مقادیر بزرگ k می‌توان براوردگر نسبتی را به جای براوردگر رگرسیونی استفاده نمود.یک راه ساده برای ترکیب (۱۰)، (۱۲) و (۱۳) این است که قرار دهیم $\rho = k_\circ \delta$، که در آن $k_\circ > ۰/۵$، $k_\circ \neq ۱$ و

$$\frac{۱-k_\circ^۲}{\left(۱-k_\circ\right)^۲}\eta_z > ۱-\bar{\alpha}_{۲(k)} .$$

تحت شرط (۱۵) براوردگر نسبتی کاراتر از براوردگر میانگین و براوردگر رگرسیونی در نمونه‌گیری از مجموعه‌ی رتبه‌دار و همچنین براوردگر نسبتی در نمونه‌گیری تصادفی ساده است.

۴-۲-۵- توزیع نرمال دومتغیره

در این بخش شرایط بیان‌شده در بخش قبل را برای معمول‌ترین توزیع دومتغیره یعنی توزیع نرمال دومتغیره بررسی می‌کنیم. در توزیع نرمال دومتغیره همان‌طورکه واترسون [۱۱] نشان داده است داریم

$$\frac{\tau_{[r]}}{\sigma_Y} = \rho \frac{\tau_{(r)}}{\sigma_X}.$$

به این ترتیب

$$\bar{\alpha}_{p,k} = \rho \bar{\alpha}_{۲(k)},$$
$$\bar{\alpha}_{۲[k]} = \rho^۲ \bar{\alpha}_{۲(k)}.$$

بنا بر این

$$\mathrm{MSE}\left(\hat{\mu}_R\right) \approx \frac{{\mu_Y}^۲}{k}\left(C_Y^۲\left(۱-\rho^۲\bar{\alpha}_{۲(k)}\right)+C_X^۲\left(۱-\bar{\alpha}_{۲(k)}\right)-۲\rho C_X C_Y\left(۱-\bar{\alpha}_{۲(k)}\right)\right).$$

در حالت $C_X = C_Y$ داریم

$$\mathrm{MSE}\left(\hat{\mu}_R\right) \approx \frac{{\sigma_Y}^۲}{k}\left(۲\left(۱-\rho\right)-\bar{\alpha}_{۲(k)}\left(۱-\rho\right)^۲\right). \qquad (۱۶)$$

برای آزمودن میزان دقت تقریب استفاده‌شده برای محاسبه‌ی میانگین توان‌های دوم خطا در حالتی که $C_X = C_Y$ و μ_X کران‌دار و دور از صفر است، مقادیر میانگین توان‌های دوم خطا با استفاده از شبیه‌سازی رایانه‌ای با استفاده از ۱۰۰۰۰ تکرار نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار از توزیع نرمال دومتغیره با پارامترهای $\mu_X = \mu_Y = ۱۰$ و $\sigma_X = \sigma_Y = ۱$ و به‌ازای مقادیر $\rho = ۰/۶(۰/۱)۰/۹, ۰/۹۵$ و $k = ۲(۲)۱۰$ تا سه رقم اعشار محاسبه و در جدول ۱ با مقادیر تقریبی محاسبه‌شده در (۱۶) مقایسه شده است.

همان‌طورکه مشاهده می‌شود اختلاف این مقادیر بسیار ناچیز است.

همچنین

$$\mathrm{B}\left(R_{k:k}\right) \approx \frac{۱}{k} \frac{\mu_Y}{\mu_X}\left(۱-\bar{\alpha}_{۲(k)}\right)\left(C_X^{۲} - \rho C_X C_Y\right).$$

مقادیر میانگین آماره‌های ترتیبی نرمال استاندارد برای $n \le ۲۰۰$ توسط پی‌یرسون و هارتلی [۶] داده شده است. برای مقادیر kی به‌اندازه‌ی کافی بزرگ $\bar{\alpha}_{۲(k)} \approx ۱$. به‌عبارت دقیق‌تر برای $n > ۵۱$، $۱ - \bar{\alpha}_{۲(k)} < ۰/۰۵$. بنا بر این با افزایش k اریبی به صفر می‌گراید. همچنین به‌ازای $\rho = \frac{C_X}{C_Y}$ اریبی تقریباً برابر صفر است. به‌خصوص اگر $C_X = C_Y$ با نزدیک شدن ρ به یک، اریبی به صفر می‌گراید. در حالت $C_X = C_Y$ داریم

$$\mathrm{B}\left(R_{k:k}\right) \approx \frac{C_X^{۲}}{k} \frac{\mu_Y}{\mu_X}\left(۱-\bar{\alpha}_{۲(k)}\right)(۱-\rho).$$

علاوه بر این با توجه به (۱۲) $\hat{\mu}_R$ از براوردگر نسبتی در یک نمونه‌ی تصادفی دومتغیره دقیق‌تر است اگر

$$۱ + \rho^{۲}\left(\frac{C_X}{C_Y}\right)^{۲} - ۲\rho \frac{C_X}{C_Y} > ۰.$$

شرط فوق به‌وضوح همواره برقرار است.

همچنین نابرابری (۵) به‌صورت زیر در می‌آید

$$\rho > \frac{C_X}{۲C_Y}\left(۱-\bar{\alpha}_{۲(k)}\right) + \rho \bar{\alpha}_{۲(k)}$$

جدول ۱– مقادیر میانگین توان‌های دوم خطای تقریبی (عدد اول هر سطر) و شبیه‌سازی‌شده (عدد دوم هر سطر) در یک توزیع نرمال دومتغیره با ضرایب تغییر حاشیه‌ای یکسان

k	ρ				
	۰/۶	**۰/۷**	**۰/۸**	**۰/۹**	**۰/۹۵**
۲	۰/۳۷۵	۰/۲۸۶	۰/۱۹۴	۰/۰۹۸	۰/۰۵۰
	۰/۳۸۷	۰/۲۹۵	۰/۱۹۷	۰/۰۹۸	۰/۰۴۹
۴	۰/۱۷۷	۰/۱۳۷	۰/۰۹۴	۰/۰۴۹	۰/۰۲۵
	۰/۱۷۴	۰/۱۳۹	۰/۰۹۵	۰/۰۴۹	۰/۰۲۵
۶	۰/۱۱۵	۰/۰۹۰	۰/۰۶۲	۰/۰۳۲	۰/۰۱۶
	۰/۱۱۶	۰/۰۹۳	۰/۰۶۲	۰/۰۳۲	۰/۰۱۶
۸	۰/۰۸۵	۰/۰۶۷	۰/۰۴۶	۰/۰۲۴	۰/۰۱۲
	۰/۰۸۵	۰/۰۶۵	۰/۰۴۷	۰/۰۲۴	۰/۰۱۲
۱۰	۰/۰۶۷	۰/۰۵۳	۰/۰۳۷	۰/۰۱۹	۰/۰۱۰
	۰/۰۶۶	۰/۰۵۳	۰/۰۳۷	۰/۰۱۹	۰/۰۱۰

جدول ۲– مقادیر کارایی نسبی تقریبی درتوزیع نرمال دومتغیره با ضرایب تغییر حاشیه‌ای یکسان

k	ρ				
	۰/۶	**۰/۷**	**۰/۸**	**۰/۹**	**۰/۹۵**
۲	۱/۱۸	۱/۴۷	۲/۰۶	۳/۷۷	۷/۱۹
۴	۱/۱۲	۱/۳۱	۱/۶۸	۲/۷۵	۴/۸۹
۶	۱/۰۹	۱/۲۳	۱/۵۱	۲/۳۰	۳/۸۸
۸	۱/۰۷	۱/۱۹	۱/۴۱	۲/۰۴	۳/۲۹
۱۰	۱/۰۶	۱/۱۶	۱/۳۴	۱/۸۷	۲/۹۲

جدول ۳- مقادیر کارایی نسبی تقریبی درتوزیع نرمال دومتغیره با ضرایب تغییر حاشیه‌ای متغیر مورد علاقه دو برابر ضریب تغییر حاشیه‌ای متغیر کمکی

k	ρ				
	۰/۶	۰/۷	۰/۸	۰/۹	۰/۹۵
۲	۱/۰۴	۱/۲۳	۱/۵۷	۲/۴۸	۳/۰۳
۴	۱/۰۲	۱/۱۴	۱/۳۶	۲/۰۷	۲/۶۲
۶	۱/۰۲	۱/۱۰	۱/۲۷	۱/۸۵	۲/۳۶
۸	۱/۰۱	۱/۰۸	۱/۲۲	۱/۷۱	۲/۱۸
۱۰	۱/۰۱	۱/۰۷	۱/۱۸	۱/۶۱	۲/۰۵

بنا بر این در توزیع نرمال دومتغیره $\hat{\mu}_R$ از $\overline{Y}_{k:k}$ دقیق‌تر است اگر

$$\rho > \frac{C_X}{۲C_Y}.$$

که همان شرط تقریبی برای ترجیح براوردگر نسبتی به براوردگر میانگین در یک نمونه‌ی تصادفی ساده‌ی دومتغیره است. علاوه بر این

$$\mathrm{RE} \approx \left(۱ + \left(\frac{C_X}{C_Y}\right)^۲ \frac{۱-\overline{\alpha}_{۲(k)}}{۱-\rho^۲\overline{\alpha}_{۲(k)}} - ۲\rho\left(\frac{C_X}{C_Y}\right)\frac{۱-\overline{\alpha}_{۲(k)}}{۱-\rho^۲\overline{\alpha}_{۲(k)}} \right)^{-۱}.$$

به خصوص در حالت $C_X = C_Y$ شرط (۵) به صورت $\rho > \frac{۱}{۲}$ در می‌آید و مقدار تقریبی کارایی نسبی برابر است با

$$\mathrm{RE} \approx \left(۱ + \left(\frac{C_X}{C_Y}\right)^۲ \frac{۱-\overline{\alpha}_{۲(k)}}{۱-\rho^۲\overline{\alpha}_{۲(k)}} - ۲\rho\left(\frac{C_X}{C_Y}\right)\frac{۱-\overline{\alpha}_{۲(k)}}{۱-\rho^۲\overline{\alpha}_{۲(k)}} \right)^{-۱}.$$

جدول ۲ مقادیر مختلف کارایی نسبی را در این حالت به ازاء مقادیر $\rho = ۰/۶(۰/۱)۰/۹, ۰/۹۵$ و $k = ۲(۲)۱۰$ نشان می‌دهد.

همان‌طور که در جدول ۲ مشاهده می‌شود، کارایی $\hat{\mu}_R$ نسبت به $\overline{Y}_{k:k}$ با کاهش k و افزایش ρ افزایش می‌یابد. همچنین همان‌طور که قبلاً نیز به آن اشاره شد، اریبی براوردگر $\hat{\mu}_R$ در حالت $C_X = C_Y$ با نزدیک شدن ρ به یک به صفر میل می‌کند. از بحث فوق می‌توان نتیجه گرفت که در توزیع نرمال دومتغیره با ضرایب تغییر یکسان زمانی که اندازه‌ی نمونه کوچک و همبستگی مثبت بالایی بین متغیر کمکی و متغیر مورد

علاقه در نمونه‌گیری به‌شیوه‌ی مجموعه‌ی رتبه‌دار وجود دارد، براوردگر نسبتی به‌مراتب بهتر از براوردگر میانگین خواهد بود.

از آن‌جا که در کاربرد براورد نسبتی شرط مثبت بودن همبستگی و میانگین‌های متغیر مورد علاقه و کمکی عموماً برقرار است، لذا کارایی نسبی بیان شده در (۱۷) ماکسیمم مقدار خود را به‌عنوان تابعی از $\frac{C_X}{C_Y}$ در $\rho = \frac{C_X}{C_Y}$ اختیار می‌کند. به این ترتیب اگرچه با کاهش این نسبت محدودیت ضریب همبستگی برای کاراتر بودن براوردگر $\hat{\mu}_R$ نسبت به $\overline{Y}_{k:k}$ کاهش می‌یابد، اما هم‌زمان میزان کارایی نسبی این براوردگر نیز برای مقادیر بزرگ‌تر کاهش خواهد یافت.

جدول ۳ برای مثال این مقادیر کارایی نسبی را به‌ازای $\frac{C_X}{C_Y} = ۰/۵$ به‌ازای $\rho = ۰/۳(۰/۲)۰/۹, ۰/۹۵$ نشان می‌دهد. همان‌طور که ملاحظه می‌کنید اگرچه شرط $\frac{C_X}{C_Y} = ۰/۵$ بیش‌ترین مقدار کارایی نسبی را برای $\rho = ۰/۵$ نتیجه می‌دهد و همچنین باعث می‌گردد تا براوردگر $\hat{\mu}_R$ به‌ازای مقادیر کوچک‌تر ρ نسبت به $\overline{Y}_{k:k}$ دقت بیش‌تری داشته باشد، اما در عین حال باعث کاهش میزان کارایی نسبی برای مقادیر $\rho > ۰/۵$ نسبت به حالت $\frac{C_X}{C_Y} = ۱$ شده است.

مرجع‌ها

[1] Arnold, B.C.; Balakrishnan, N.; Nagaraja, H.N. (1992). *A First Course in Order Statistics.* Wiley, New York.

[2] Dell, T.R.; Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, **28**, 545-555.

[3] Kadilar, C.; Unyazici, Y.; Cingi, H. (2007). Ratio estimator for the population mean using ranked set sampling. *Stat. papers*, **50**, 301-309.

[4] McIntyre, G.A. (1952). A method of unbiased selective sampling, using ranked sets. *J. Agri. Res.* **3**, 385-390.

[5] Pearson, E.S.; Hartley, H.O. (1972). *Biometrika Tables for Statisticians,* Vol. 2. Cambridge University Press, England.

[6] Samawi, H.M.; Muttlak, H.A. (1996). Estimation of ratio using rank set sampling. *Biometr J,* **38**, 753-764.

[7] Samawi, H.M.; Al-Saleh; Fraiwan, M. (2004). On bivariate ranked set sampling for distribution and quantile estimation and quantile interval estimation using ratio estimator. *Comm. Statist. Theory Methods,* **33**, 1801-1819.

[8] Stokes, S.L. (1980a). Estimation of variance using judgment ordered ranked set samples. *Biometrics,* **36**, 35-42.

[9] Stokes, S. L. (1980b). Inference on the correlation coefficient in bivariate normal populations from ranked set samples. *J. Amer. Statist. Assos,* **75**, 989-995.

[10] Watterson, G.A. (1959). Linear estimation in censored samples from multivariate normal populations. *Ann. Math. Statist.* **30**, 814-824.

[11] Yu, Philip, L.H.; Lam, K. (1997). Regression estimator in ranked set sampling. *Biometrics,* **53**, 1070-1080.

مرتضی امینی
دانشجوی دکترای آمار
مشهد، میدان آزادی، دانشگاه فردوسی مشهد، دانشکده‌ی علوم ریاضی، گروه آمار.
پیام‌نگار: mort.amini@gmail.com

ناصررضا ارقامی
استاد آمار
مشهد، میدان آزادی، دانشگاه فردوسی مشهد، دانشکده‌ی علوم ریاضی، گروه آمار.
پیام‌نگار: nr_arghami@yahoo.com