

برآورد قابلیت اعتماد پویا برای تنش و مقاومت وابسته بر اساس نمونه گیری مجموعه رتبه دار

مهدی مهدی زاده^{۱*} و احسان زمان زاده^{**}

^{*}گروه آمار، دانشگاه حکیم سبزواری

^{**}گروه آمار، دانشگاه اصفهان

تاریخ دریافت: ۱۳۹۶/۴/۱۵ تاریخ پذیرش: ۱۳۹۷/۵/۲۶

چکیده: در این مقاله، برآورد یک معیار قابلیت اعتماد پویا بر اساس نمونه گیری مجموعه رتبه دار مطالعه می شود وقتی که متغیرهای تنش و مقاومت وابسته اند. برآوردگر پیشنهادی با برآوردگر متناظر در نمونه گیری تصادفی ساده به صورت نظری و عددی مقایسه شده است. نتایج نشان می دهد برآورد حاصل از طرح نمونه گیری مجموعه رتبه دار غالباً دقیق تر است. همچنین، تفاوت دقت دو برآوردگر با افزایش ضریب همبستگی بین متغیرهای تنش و مقاومت محسوس تر می شود.

واژه های کلیدی: رتبه بندی قضاوتی؛ قابلیت اعتماد؛ نمونه گیری مجموعه رتبه دار.

رده بندی موضوعی (۲۰۱۰): ۶۲G۳۰، ۶۲N۰۵.

۱- مقدمه

طرح نمونه گیری مجموعه رتبه دار^۲ (RSS) توسط مک اینتایر [۱] برای بهبود برآورد میانگین محصول مرتع پیشنهاد شد. این روش برای شرایطی مناسب است که اندازه گیری متغیر مورد نظر دشوار (هزینه بر، وقت گیر یا مخرب) است اما واحدهای نمونه گیری را می توان نسبتاً آسان و با هزینه کم بر اساس صفت مورد نظر رتبه بندی کرد. رتبه بندی به صورت قضاوتی^۳ (بدون اندازه گیری دقیق) و معمولاً با استفاده از یک متغیر همراه (که با متغیر اصلی همبستگی بالا دارد)، یا نظر یک کارشناس انجام می شود.

۱- آدرس الکترونیکی نویسنده مسئول مقاله: mahdizadeh.m@live.com

2- Ranked set sampling

3- Judgmental

مراحل این طرح عبارت‌اند از:

۱. نمونه‌ای تصادفی به حجم m^2 از جامعه در نظر گرفته و آن را به تصادف به m مجموعه هر یک به حجم m تخصیص می‌دهیم.

۲. اعضای هر مجموعه با رتبه‌بندی قضاوتی نسبت به مشخصه موردنظر مرتب‌شده و عضو با رتبه i ($i = 1, \dots, m$) در مجموعه i ام، برای اندازه‌گیری دقیق انتخاب می‌شود.

برای کاهش خطای رتبه‌بندی، m باید کوچک باشد. بنابراین، رویه فوق می‌تواند l مرتبه (چرخه^۱) تکرار شود تا یک نمونه مجموعه رتبه‌دار به حجم mm به دست آید. با فرض حجم نمونه کل یکسان، روش‌های آماری در این طرح از معادل خود در نمونه‌گیری تصادفی ساده^۲ (SRS) معمولاً کاراترند. برای مرور تاریخچه و کاربردهای این طرح نمونه‌گیری به چن و همکاران [۲] مراجعه کنید.

گونه‌های زیادی از RSS توسعه یافته‌اند که هر یک تحت شرایطی به بهبود روش‌های استنباطی منجر می‌شوند. در ادامه، یک تعمیم دومتغیره از آن ارائه می‌شود که جایگزینی برای نمونه‌گیری تصادفی ساده دومتغیره^۳ (BVSRS) است. فرض کنید (X, Y) یک بردار تصادفی دومتغیره با تابع چگالی احتمال $f_{X,Y}(x, y)$ باشد. طرح نمونه‌گیری مجموعه رتبه‌دار دومتغیره^۴ (BVRSS) که توسط الصالح و زنگ [۳] معرفی شد، به صورت زیر انجام می‌شود:

۱. نمونه‌ای تصادفی به حجم m^4 از جامعه در نظر گرفته و آن را به تصادف به m^2 گروه هر یک به حجم m^2 تخصیص می‌دهیم که اعضای هر گروه، آرایشی مانند یک ماتریس $m \times m$ دارند.

۲. در گروه اول، اعضای هر سطر با رتبه‌بندی قضاوتی نسبت به مشخصه اول (X) مرتب‌شده و اعضای با رتبه یک تعیین می‌شوند.

-
- 1- Cycle
 - 2- Simple random sampling
 - 3- Bivariate simple random sampling
 - 4- Bivariate ranked set sampling

۳. m عضو حاصل از گام ۲، با رتبه‌بندی قضاوتی نسبت به مشخصه دوم (Y) مرتب‌شده و عضو با رتبه یک برای اندازه‌گیری دقیق انتخاب می‌شود. این عضو را با برچسب $(1,1)$ نشان می‌دهیم.

۴. گام‌های ۲ و ۳ در مورد گروه دوم تکرار شده با این تفاوت که در گام ۳، عضو با رتبه دو برای اندازه‌گیری دقیق انتخاب می‌شود. این عضو را با برچسب $(1,2)$ نشان می‌دهیم.

۵. این فرایند ادامه می‌یابد تا وقتی که از گروه آخر (m^2)، عضو با برچسب (m,m) انتخاب شود.

با r چرخه تکرار رویه فوق، یک نمونه مجموعه رتبه‌دار دومتغیره به حجم $n = m^2$ استخراج می‌شود. نمونه حاصل را با $\left\{ \left(X_{[i](j)}^k, Y_{(i)[j]}^k \right) : i, j = 1, \dots, m; k = 1, \dots, r \right\}$ نشان می‌دهیم که در آن $\left(X_{[i](j)}^k, Y_{(i)[j]}^k \right)$ عضو با برچسب (i, j) در چرخه k ام است. فرض کنید $f_{X_{[i](j)}, Y_{(i)[j]}}(x, y)$ تابع چگالی احتمال $\left(X_{[i](j)}, Y_{(i)[j]} \right)$ باشد. آنگاه می‌توان دید که اتحاد

$$f(x, y) = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m f_{X_{[i](j)}, Y_{(i)[j]}}(x, y) \quad (1)$$

برقرار است (برای اثبات، لم ۲ در الصالح و ژنگ [۳] را ببینید). این نتیجه، نقش مهمی در مطالعه رویه‌های آماری بر اساس BVRSS دارد. برای مثال، الصالح و سماوی [۴] از این نتیجه برای برآورد ضریب همبستگی بین دو متغیر استفاده کردند.

استنباط آماری در مورد پارامتر $\theta = P(X > Y)$ وقتی X و Y دو متغیر تصادفی هستند، موضوعی مهم در قابلیت اعتماد بوده که مورد توجه محققان زیادی قرار گرفته است. اگر X میزان مقاومت یک سیستم، و Y میزان تنش وارد شده به آن باشد، سیستم در صورتی به کار خود ادامه می‌دهد که مقاومت آن از تنش وارد شده بیشتر باشد. بنابراین، پارامتر θ احتمال کار کردن سیستم را نشان می‌دهد. علاوه بر مبحث قابلیت اعتماد، برآورد این پارامتر در پزشکی، روانشناسی، علوم تربیتی و ... نیز کاربرد دارد (کوئز و همکاران [۵]). به‌عنوان مثال، اگر X و Y به ترتیب طول عمر بیماران باشد که از داروی جدید و دارونما استفاده می‌کنند، آنگاه پارامتر θ را می‌توان به‌عنوان معیار اثربخشی داروی جدید تفسیر کرد. شایان ذکر است که مطالعاتی

درباره برآورد کمیت فوق در RSS و BVRSS انجام شده است (به سن‌گویتا و موختی [۶] و سماوی و همکاران [۷] مراجعه کنید).

زاردشت و اسدی [۸] یک شاخص قابلیت اعتماد پویا^۱ (وابسته به زمان) به صورت

$$\theta(t) = \frac{\alpha(t)}{\beta(t)}$$

معرفی کردند که در آن $\alpha(t) = P(X > Y > t)$ و $\beta(t) = P(X > t, Y > t)$. اگر X و Y طول عمر دو سیستم باشند، آنگاه $\theta(t)$ احتمال آن است که طول عمر سیستم اول از سیستم دوم بیشتر باشد، با فرض این که هر دو سیستم تا زمان t کار کرده‌اند. به‌عنوان مثال، فرض کنید X و Y به ترتیب طول عمر بیماران مبتلا به نوعی تومور سرطانی باشد که با روش جراحی باز و روش لیزر عمل شده‌اند. در این صورت، $\theta(t)$ را می‌توان به‌عنوان شاخص بهتر بودن روش اول از روش دوم، بعد از گذشت زمان t از انجام عمل در نظر گرفت.

زاردشت و اسدی [۸] برآورد $\theta(t)$ را وقتی X و Y دو متغیر تصادفی مستقل هستند، تحت SRS مطالعه کردند. مهدی زاده و زمان زاده [۹] برآوردگر مشابهی تحت RSS ارائه کردند. در این مقاله قصد داریم برآورد $\theta(t)$ را در حالتی که دو متغیر تصادفی وابسته‌اند بررسی کنیم.

در بخش ۲، برآوردگرهای پیشنهادی تحت BVRSS و BVSRS معرفی می‌شوند. عملکرد این برآوردگرها با استفاده از شبیه‌سازی مونت‌کارلو ارزیابی شده که نتایج آن در بخش ۳ آمده است. در بخش ۴، کاربرد رویه پیشنهادی را به کمک یک مجموعه داده پزشکی توضیح می‌دهیم. یافته‌های مقاله و آینده تحقیق در بخش ۵ آمده است.

۲- برآوردگر قابلیت اعتماد پویا

نمونه مجموعه رتبه‌دار $\left\{ \left(X_{[i]_{(j)}}^k, Y_{(i)_{[j]}}^k \right) : i, j = 1, \dots, m; k = 1, \dots, r \right\}$ استخراج شده از $f_{X,Y}(x,y)$ را در نظر بگیرید که در آن $n = rm^2$. فرض کنید می‌خواهیم $\theta(t)$ را بر اساس این نمونه برآورد کنیم. با استفاده از تابع توزیع تجربی، برآوردگر

$$\tilde{\theta}(t) = \frac{\tilde{\alpha}(t)}{\tilde{\beta}(t)} \quad (۲)$$

پیشنهاد می‌شود که در آن

$$\tilde{\alpha}(t) = \frac{1}{n} \sum_{k=1}^r \sum_{i=1}^m \sum_{j=1}^m I(X_{[i](j)}^k > Y_{(i)[j]}^k > t),$$

$$\tilde{\beta}(t) = \frac{1}{n} \sum_{k=1}^r \sum_{i=1}^m \sum_{j=1}^m I(X_{[i](j)}^k > t, Y_{(i)[j]}^k > t).$$

معادل برآوردگر (۲) بر اساس نمونه تصادفی ساده $\{(X_i, Y_i) : i = 1, \dots, n\}$ استخراج شده از $f_{X,Y}(x,y)$ عبارت است از

$$\hat{\theta}(t) = \frac{\hat{\alpha}(t)}{\hat{\beta}(t)} \quad (3)$$

که در آن

$$\hat{\alpha}(t) = \frac{1}{n} \sum_{i=1}^n I(X_i > Y_i > t),$$

$$\hat{\beta}(t) = \frac{1}{n} \sum_{i=1}^n I(X_i > t, Y_i > t).$$

می‌توان دید $\hat{\alpha}(t)$ و $\hat{\beta}(t)$ برآوردگرهای ناریب برای $\alpha(t)$ و $\beta(t)$ بوده و واریانس آن‌ها به ترتیب برابر است با

$$\text{Var}(\hat{\alpha}(t)) = \frac{\alpha(t)(1-\alpha(t))}{n},$$

$$\text{Var}(\hat{\beta}(t)) = \frac{\beta(t)(1-\beta(t))}{n}.$$

در ادامه، ویژگی‌های این دو برآوردگر بررسی می‌شود. به این منظور، ابتدا کمیت‌های $\alpha_{i,j}(t)$ و $\beta_{i,j}(t)$ را به صورت

$$\alpha_{i,j}(t) = P(X_{[i](j)}^1 > Y_{(i)[j]}^1 > t)$$

$$\beta_{i,j}(t) = P(X_{[i](j)}^1 > t, Y_{(i)[j]}^1 > t)$$

تعریف می‌کنیم.

قضیه ۱: برای هر t ثابت، برآوردگرهای $\tilde{\alpha}(t)$ و $\tilde{\beta}(t)$ نارایب هستند.

برهان: با استفاده از رابطه (۱) داریم

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^m \alpha_{i,j}(t) &= \sum_{i=1}^m \sum_{j=1}^m P(X_{[i](j)} > Y_{(i)[j]} > t) \\ &= \sum_{i=1}^m \sum_{j=1}^m \iint_{x>y>t} f_{X_{[i](j)} Y_{(i)[j]}}(x,y) dx dy \\ &= m^\nu \iint_{x>y>t} f_{X,Y}(x,y) dx dy \\ &= m^\nu \alpha(t). \end{aligned} \quad (۴)$$

به‌طور مشابه می‌توان نشان داد

$$\sum_{i=1}^m \sum_{j=1}^m \beta_{i,j}(t) = m^\nu \beta(t). \quad (۵)$$

اثبات با توجه به روابط (۴) و (۵) کامل است. ■

لم زیر (که به‌سادگی اثبات می‌شود) در اثبات قضیه بعدی به کار می‌رود.

لم ۱: برای $\alpha_{i,j}(t)$ ، $\alpha(t)$ ، $\beta_{i,j}(t)$ و $\beta(t)$ معرفی‌شده در بالا داریم

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^m \alpha_{i,j}(t)(1 - \alpha_{i,j}(t)) &= m^\nu \alpha(t)(1 - \alpha(t)) - \sum_{i=1}^m \sum_{j=1}^m (\alpha_{i,j}(t) - \alpha(t))^2, \\ \sum_{i=1}^m \sum_{j=1}^m \beta_{i,j}(t)(1 - \beta_{i,j}(t)) &= m^\nu \beta(t)(1 - \beta(t)) - \sum_{i=1}^m \sum_{j=1}^m (\beta_{i,j}(t) - \beta(t))^2. \end{aligned}$$

اکنون با استفاده از لم فوق نشان می‌دهیم برآوردگرهای $\alpha(t)$ و $\beta(t)$ در BVRSS نسبت به معادل خود در BVSRS واریانس کمتری دارند.

قضیه ۲: برای هر t ثابت، $\text{Var}(\tilde{\alpha}(t)) \leq \text{Var}(\hat{\alpha}(t))$ و $\text{Var}(\tilde{\beta}(t)) \leq \text{Var}(\hat{\beta}(t))$.

برهان: با استفاده از لم ۱ می‌توان نوشت

$$\begin{aligned}
\text{Var}(\tilde{\alpha}(t)) &= \frac{r}{n^{\gamma}} \sum_{i=1}^m \sum_{j=1}^m \text{Var}\left(I(X_{[i](j)}^{\downarrow} > Y_{(i)[j]}^{\downarrow} > t)\right) \\
&= \frac{r}{n^{\gamma}} \sum_{i=1}^m \sum_{j=1}^m \left(\alpha_{i,j}(t) - \alpha_{i,j}^{\gamma}(t)\right) \\
&= \frac{r}{n^{\gamma}} \left[m^{\gamma} \alpha(t)(1 - \alpha(t)) - \sum_{i=1}^m \sum_{j=1}^m \left(\alpha_{i,j}(t) - \alpha(t)\right)^{\gamma} \right] \\
&= \frac{\alpha(t)(1 - \alpha(t))}{n} - \frac{1}{nm^{\gamma}} \sum_{i=1}^m \sum_{j=1}^m \left(\alpha_{i,j}(t) - \alpha(t)\right)^{\gamma} \\
&\leq \text{Var}(\hat{\alpha}(t)).
\end{aligned}$$

به‌طور مشابه می‌توان نشان داد

$$\text{Var}(\tilde{\beta}(t)) \leq \text{Var}(\hat{\beta}(t)). \blacksquare$$

با استفاده از قضیه نگاشت پیوسته (شانو [۱۰]، ص. ۵۹) نتیجه می‌شود که $\hat{\theta}(t)$ و $\tilde{\theta}(t)$ قویاً سازگارند. این ویژگی و قضیه همگرایی مغلوب تضمین می‌کنند که این دو برآوردگر به‌طور مجانبی ناریب هستند. کمیت‌های لازم برای تقریب میانگین مربع خطای این برآوردگرها در قضیه زیر آمده است.

قضیه ۳: کوواریانس بین $\hat{\alpha}(t)$ و $\hat{\beta}(t)$ ، و کوواریانس بین $\tilde{\alpha}(t)$ و $\tilde{\beta}(t)$ به ترتیب عبارت‌اند از

$$\text{Cov}(\hat{\alpha}(t), \hat{\beta}(t)) = \frac{1}{n} \alpha(t)(1 - \beta(t)), \quad (۶)$$

$$\text{Cov}(\tilde{\alpha}(t), \tilde{\beta}(t)) = \frac{1}{nm^{\gamma}} \left[\sum_{i=1}^m \sum_{j=1}^m \alpha_{i,j}(t)(1 - \beta_{i,j}(t)) \right]. \quad (۷)$$

برهان: برای اثبات رابطه (۶)، با استفاده از تعریف $\hat{\alpha}(t)$ و $\hat{\beta}(t)$ می‌توان نوشت

$$\begin{aligned} \text{Cov}(\hat{\alpha}(t), \hat{\beta}(t)) &= \frac{1}{n^{\gamma}} \text{Cov} \left(\sum_{i=1}^n I(X_i > Y_i > t), \sum_{j=1}^n I(X_j > t, Y_j > t) \right) \\ &= \frac{1}{n^{\gamma}} \sum_{i=1}^n \text{Cov}(I(X_i > Y_i > t), I(X_i > t, Y_i > t)) \\ &= \frac{1}{n} \alpha(t)(1 - \beta(t)). \end{aligned}$$

به‌طور مشابه، برای اثبات رابطه (۷) داریم

$$\begin{aligned} \text{Cov}(\tilde{\alpha}(t), \tilde{\beta}(t)) &= \frac{1}{n^{\gamma}} \text{Cov} \left(\sum_{k=1}^r \sum_{i=1}^m \sum_{j=1}^m I(X_{[i](j)}^k > Y_{(i)[j]}^k > t), \right. \\ &\quad \left. \sum_{k'=1}^r \sum_{i'=1}^m \sum_{j'=1}^m I(X_{[i'](j')}^{k'} > t, Y_{(i')[j']}^{k'} > t) \right) \\ &= \frac{r}{n^{\gamma}} \sum_{i=1}^m \sum_{j=1}^m \text{Cov} \left(I(X_{[i](j)}^1 > Y_{(i)[j]}^1 > t), I(X_{[i](j)}^1 > t, Y_{(i)[j]}^1 > t) \right) \\ &= \frac{1}{nm^{\gamma}} \left[\sum_{i=1}^m \sum_{j=1}^m \alpha_{i,j}(t)(1 - \beta_{i,j}(t)) \right]. \end{aligned}$$

با استفاده از بسط تیلور مرتبه یک می‌توان نوشت

$$\hat{\theta}(t) - \theta(t) \approx \frac{1}{\beta(t)} (\hat{\alpha}(t) - \alpha(t)) - \frac{\alpha(t)}{\beta^{\gamma}(t)} (\hat{\beta}(t) - \beta(t)), \quad (۸)$$

$$\tilde{\theta}(t) - \theta(t) \approx \frac{1}{\beta(t)} (\tilde{\alpha}(t) - \alpha(t)) - \frac{\alpha(t)}{\beta^{\gamma}(t)} (\tilde{\beta}(t) - \beta(t)). \quad (۹)$$

بر اساس دو رابطه فوق، تقریب میانگین مربع خطای $\hat{\theta}(t)$ و $\tilde{\theta}(t)$ به ترتیب عبارت‌اند از

$$\text{MSE}_{\text{appr.}}(\hat{\theta}(t)) = \frac{1}{\beta^{\gamma}(t)} \text{Var}(\hat{\alpha}(t)) + \frac{\alpha^{\gamma}(t)}{\beta^{\gamma}(t)} \text{Var}(\hat{\beta}(t)) - \gamma \frac{\alpha(t)}{\beta^{\gamma}(t)} \text{Cov}(\hat{\alpha}(t), \hat{\beta}(t)), \quad (۱۰)$$

$$\text{MSE}_{\text{appr.}}(\tilde{\theta}(t)) = \frac{1}{\beta^{\gamma}(t)} \text{Var}(\tilde{\alpha}(t)) + \frac{\alpha^{\gamma}(t)}{\beta^{\gamma}(t)} \text{Var}(\tilde{\beta}(t)) - \gamma \frac{\alpha(t)}{\beta^{\gamma}(t)} \text{Cov}(\tilde{\alpha}(t), \tilde{\beta}(t)). \quad (۱۱)$$

با توجه به قضیه ۲ می‌دانیم نامساوی‌های $\text{Var}(\tilde{\alpha}(t)) \leq \text{Var}(\hat{\alpha}(t))$ و $\text{Var}(\tilde{\beta}(t)) \leq \text{Var}(\hat{\beta}(t))$ برقرار است. اکنون اگر بتوان نشان داد

$$\text{Cov}(\hat{\alpha}(t), \hat{\beta}(t)) \leq \text{Cov}(\tilde{\alpha}(t), \tilde{\beta}(t)), \quad (12)$$

آنگاه از روابط (۱۰) و (۱۱) نتیجه می‌شود برای هر t ثابت، میانگین مربع خطای تقریبی $\tilde{\theta}(t)$ از مقدار مشابه برای $\hat{\theta}(t)$ کمتر است. اما با توجه به شکل عبارات حاصل، اثبات رابطه (۱۲) آسان به نظر نمی‌رسد.

این بخش را با یک قضیه در مورد توزیع مجانبی برآوردگرهای موردبررسی به پایان می‌بریم.

قضیه ۴: اگر در طرح‌های BVSRS و BVRSS به ترتیب $n \rightarrow \infty$ و $r \rightarrow \infty$ ، آنگاه

$$\sqrt{n}(\hat{\theta}(t) - \theta(t)) \xrightarrow{d} N\left(0, n \text{MSE}_{\text{appr.}}(\hat{\theta}(t))\right)$$

$$\sqrt{n}(\tilde{\theta}(t) - \theta(t)) \xrightarrow{d} N\left(0, n \text{MSE}_{\text{appr.}}(\tilde{\theta}(t))\right),$$

که در آن $\text{MSE}_{\text{appr.}}(\hat{\theta}(t))$ و $\text{MSE}_{\text{appr.}}(\tilde{\theta}(t))$ به ترتیب در رابطه (۱۰) و (۱۱) آمده‌اند.

برهان: با استفاده از رابطه (۸) می‌توان دید $\hat{\theta}(t) - \theta(t) = \sum_{i=1}^n W_i / n + U_n$ که در آن

$U_n \xrightarrow{p} 0$ و W_i ($i = 1, \dots, n$) متغیرهای تصادفی مستقل و هم توزیع هستند و

$$W_i = \frac{1}{\beta(t)} I(X_i > Y_i > t) - \frac{\alpha(t)}{\beta^r(t)} I(X_i > t, Y_i > t).$$

به‌طور مشابه، از رابطه (۹) نتیجه می‌شود $\tilde{\theta}(t) - \theta(t) = \sum_{k=1}^r \sum_{i=1}^m \sum_{j=1}^m Z_{i,j}^k / n + V_r$ که در آن

$V_r \xrightarrow{p} 0$ و برای i و j ثابت، $Z_{i,j}^k$ ($k = 1, \dots, r$) متغیرهای تصادفی مستقل و هم توزیع هستند و

$$Z_{i,j}^k = \frac{1}{\beta(t)} I(X_{[i](j)}^k > Y_{(i)[j]}^k > t) - \frac{\alpha(t)}{\beta^r(t)} I(X_{[i](j)}^k > t, Y_{(i)[j]}^k > t).$$

■ اکنون با استفاده از قضیه‌های حد مرکزی و اسلاتسکی، اثبات کامل می‌شود.

۳- نتایج عددی

در این بخش، از شبیه‌سازی مونت کارلو برای مقایسه میانگین مربع خطای $\hat{\theta}(t)$ و $\tilde{\theta}(t)$ استفاده می‌شود. به این منظور، ابتدا فرض می‌کنیم (X, Y) توزیع نرمال دومتغیره با تابع چگالی احتمال

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi\sigma_1\sigma_2}\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x-\mu_1}{\sigma_1} - \rho\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)\right]^2\right\}$$

داشته باشند که آن را به اختصار با $(X, Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ نشان می‌دهیم. در اینجا، حالتی را در نظر می‌گیریم که $(X, Y) \sim N_2(0, 0, 1, 1, 1/5)$ و $\rho = 0.1, 0.5, 0.9$ باشد.

برای در نظر گرفتن توزیع‌های دومتغیره دیگر، ابتدا مقدماتی درباره تابع مفصل^۱ بیان می‌کنیم. تابع مفصل، تابعی است که توابع توزیع یک متغیره را برای تشکیل یک تابع توزیع چند متغیره به هم پیوند می‌دهد. مزیت استفاده از مفصل برای تولید توزیع‌های چند متغیره این است که مفصل می‌تواند اثرات وابستگی را از اثرات توزیع‌های حاشیه‌ای مجزا کند.

برای دو متغیر تصادفی پیوسته X و Y به ترتیب با تابع توزیع $F(x)$ و $G(y)$ ، و تابع توزیع توأم $H(x, y)$ ، تابع مفصلی مانند $C(\cdot, \cdot)$ وجود دارد به طوری که

$$H(x, y) = C(F(x), G(y)), \quad \forall x, y \in \mathbb{R}.$$

اگر $F(x)$ و $G(y)$ مطلقاً پیوسته باشند، آنگاه تابع مفصل C یکتاست (اسکلار [۱۱]). فرض کنید $f(x)$ و $g(y)$ به ترتیب تابع چگالی حاشیه‌ای X و Y باشد. در این صورت، تابع چگالی توأم آن‌ها عبارت است از

$$h(x, y) = c(F(x), G(y))f(x)g(y), \quad \forall x, y \in \mathbb{R},$$

که در آن $c(F(x), G(y)) = \frac{\partial^2 C(F(x), G(y))}{\partial F(x) \partial G(y)}$ چگالی تابع مفصل است. مفصل فرانک^۲ یکی از اعضای خانواده مفصل ارشمیدسی^۳ است که چگالی تابع مفصل آن به شکل

$$c(F(x), G(y)) = \frac{\lambda(1 - \exp\{-\lambda\}) \exp\{-\lambda(F(x) + G(y))\}}{[1 - \exp\{-\lambda\} - (1 - \exp\{-\lambda F(x)\})(1 - \exp\{-\lambda G(y)\})]^2}$$

- 1- Copula function
- 2- Frank copula
- 3- Archimedean copula

است که در آن $\lambda \in \mathbb{R} \setminus \{0\}$. اکنون با استفاده از مفصل فرانک، دو توزیع دومتغیره دیگر برای استفاده در مطالعه شبیه‌سازی انتخاب می‌کنیم. فرض کنید متغیر تصادفی X که توزیع نمایی با تابع چگالی

$$f(x) = \gamma \exp\{-\gamma x\}, \quad x > 0; \gamma > 0$$

دارد را با نماد $X \sim E(\gamma)$ نشان دهیم. توزیع دومتغیره نخست، با انتخاب $(1) X \sim E(1)$ ، $Y \sim E(2)$ و $\lambda = 0.8, 4/75, 29$ ساخته می‌شود. این مقادیر λ به ترتیب به ضریب همبستگی پیرسن $\rho = 0.1, 0.5, 0.9$ بین دو متغیر منجر می‌شود. اگر متغیر تصادفی X با توزیع گاما و تابع چگالی

$$f(x) = \frac{1}{\Gamma(a)b^a} x^{a-1} \exp\left\{-\frac{x}{b}\right\}, \quad x > 0; a, b > 0$$

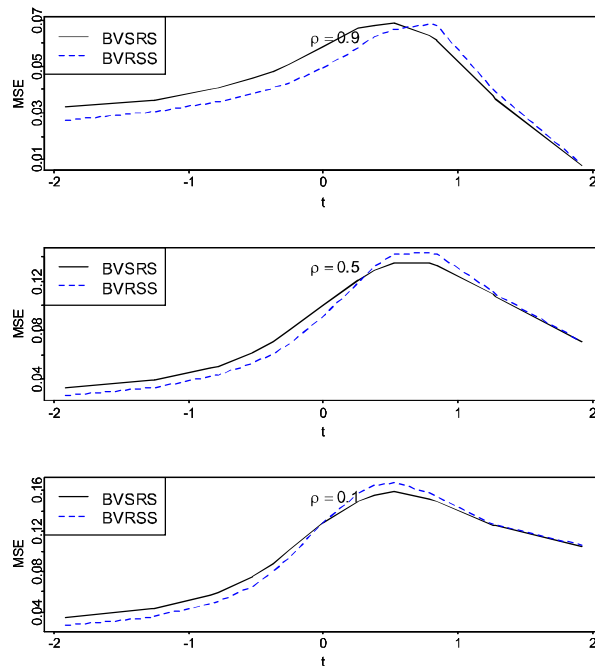
را با نماد $X \sim Ga(a, b)$ نشان دهیم، آنگاه توزیع دومتغیره بعدی، با انتخاب $(2) X \sim Ga(2, 1)$ ، $Y \sim Ga(2, 1/1)$ و $\lambda = 0.72, 4/2, 22$ ساخته می‌شود. مجدداً، با انتخاب این مقادیر λ ، ضریب همبستگی پیرسن بین دو متغیر به ترتیب $\rho = 0.1, 0.5, 0.9$ خواهد بود.

برای هر مقدار t ، میانگین مربع خطای دو برآوردگر را بر اساس $10,000$ نمونه به حجم $n = 9, 25$ تولیدشده در طرح‌های BVSRS و BVRSS برآورد می‌کنیم. در استخراج نمونه مجموعه رتبه‌دار دومتغیره، $m = 3, 5$ و $r = 1$ استفاده شد. نتایج حاصل برای سه توزیع انتخاب‌شده و مقادیر مختلف ضریب همبستگی بین دو متغیر در شکل‌های ۶-۱ رسم شده است. ملاحظه می‌شود که برآوردگر (۲) غالباً عملکرد بهتری نسبت به برآوردگر (۳) دارد. این ویژگی گاهی برای بازه‌ای از مقادیر t نقض می‌شود. برای n ثابت، میانگین مربع خطای هر برآوردگر با افزایش ρ کاهش می‌یابد. همان‌طور که انتظار داریم با فرض ثابت بودن ρ ، بزرگ‌تر شدن n باعث افزایش دقت هر برآوردگر می‌شود. به‌طور کلی، روند نمودارهای متناظر برای $n = 9$ و $n = 25$ تشابه زیادی دارند. با فرض ثابت بودن سایر عوامل، تفاوت دقت دو برآوردگر با افزایش ρ محسوس‌تر می‌شود.

۴- کاربرد

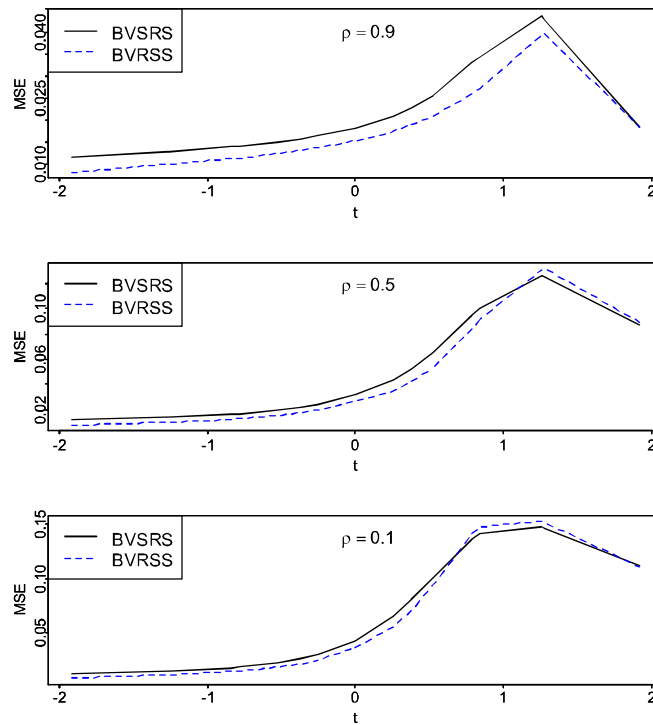
در این بخش، کاربرد برآوردگرهای پیشنهادی را با استفاده از یک مجموعه داده پزشکی توضیح می‌دهیم. در سال‌های اخیر، محققان علوم پزشکی دریافته‌اند آپولیپو پروتئین‌های انتقال‌دهنده چربی اهمیت ویژه‌ای دارند. به‌عنوان مثال، می‌توان به apoA و apoB اشاره کرد که به ترتیب

انتقال‌دهنده کلسترول خوب^۱ (HDL) و کلسترول بد^۲ (LDL) هستند (والدیوس و همکاران [۱۲]). در افراد سالم، مقادیر apoA از apoB معمولاً بیشتر است و در نتیجه آن‌ها کمتر در معرض بیماری‌های قلبی عروقی هستند.



شکل (۱): برآورد میانگین مربع خطای $\hat{\theta}(t)$ و $\tilde{\theta}(t)$ تحت توزیع نرمال دومتغیره برای $n = 9$. این آپولیپو پروتئین‌ها می‌توانند به‌عنوان جایگزین نشانگرهای^۳ زیستی رایج نظیر HDL و LDL به کار روند. در مقایسه با نشانگرهای HDL-C و LDL-C، اندازه‌گیری شاخص apoB/apoA-I (نسبت apoB به apoA-I) نیاز به ناشتا بودن ندارد که خود یک مزیت محسوب می‌شود. همچنین، مقادیر apoA-I و apoB حاصل از مطالعات مختلف، قابل مقایسه هستند (والدیوس و جانگنر [۱۳]).

- 1- High-density lipoprotein
- 2- Low-density lipoprotein
- 3- Biomarkers

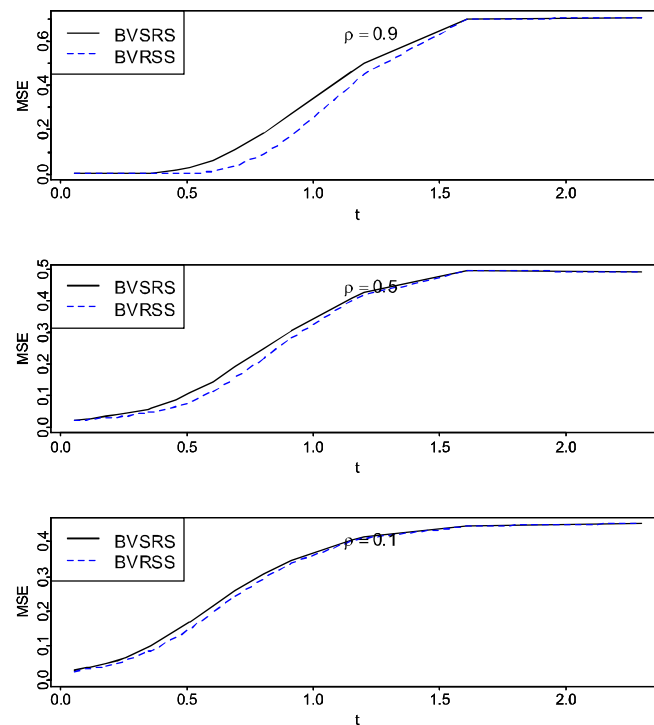


شکل (۲): برآورد میانگین مربع خطای $\hat{\theta}(t)$ و $\bar{\theta}(t)$ تحت توزیع نرمال دومتغیره برای $n = ۲۵$.

به‌جای اندازه‌گیری apoB/apoA-I، می‌توان از شاخص $\theta = P(\text{apoA} > \text{apoB})$ استفاده کرد. اگر این احتمال به‌طور معناداری بزرگ‌تر از $۰/۵$ باشد، آنگاه نتیجه می‌شود که جامعه تحت بررسی، کمتر در معرض بیماری‌های قلبی عروقی قرار دارد. اکنون فرض کنید بر مبنای مطالعات پیشین، پژوهشگر می‌داند که یک کران پایین مانند t برای مقادیر نشانگرهای زیستی در جامعه وجود دارد. در این صورت، بهتر است که شاخص

$$\theta(t) = \frac{P(\text{apoA} > \text{apoB} > t)}{P(\text{apoA} > t, \text{apoB} > t)}$$

را به کاربرد.

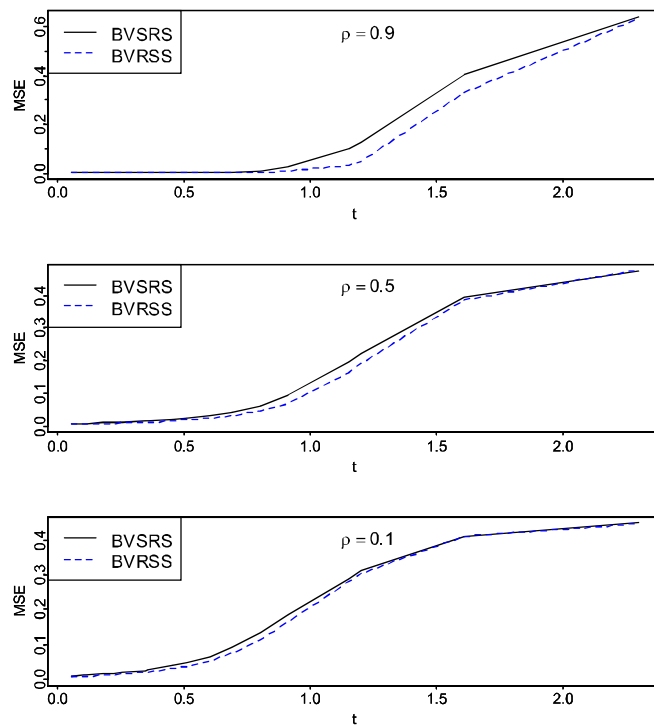


شکل (۳): برآورد میانگین مربع خطای $\hat{\theta}(t)$ و $\tilde{\theta}(t)$ تحت توزیع نمایی دومتغیره برای $n = 9$.

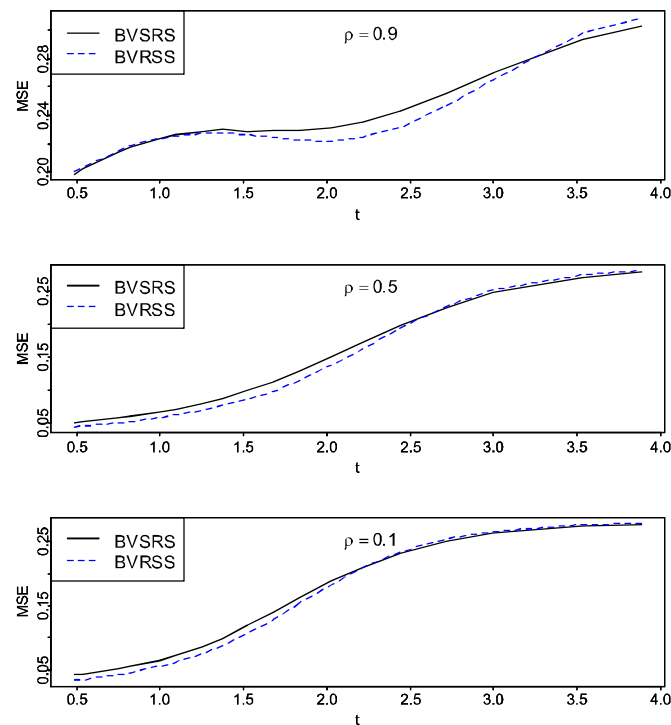
مجموعه داده CHNS^۱، مقادیر نشانگرهای زیستی apoB و apoA را برای ۱۰۱۸۷ نفر (با سن ۷ سال یا بیشتر) در سال ۲۰۰۹ گزارش می‌کند. سماوی و همکاران [۷] از این مجموعه داده برای نشان دادن کاربرد مسئله برآورد $\theta = P(X > Y)$ استفاده کردند. با در نظر گرفتن این مجموعه داده به‌عنوان جامعه تحت مطالعه، می‌خواهیم میانگین مربع خطای برآوردگرهای (۲) و (۳) را در برآورد پارامتر $\theta(0/3) = 0.765$ مقایسه کنیم. به این منظور، ۱۰۰۰ نمونه به حجم $n = 25$ در هر یک از طرح‌های BVSRS و BVRSS تولید کرده و از هر نمونه، برآورد پارامتر موردنظر را به دست می‌آوریم (در تولید نمونه مجموعه رتبه‌دار دومتغیره، $m = 5$ و $r = 1$ انتخاب شد). با استفاده از مقادیر حاصل، نسبت برآورد میانگین مربع خطای $\hat{\theta}(0/3)$ به کمیت

1- <http://www.cpc.unc.edu/projects/china/data/datasets>

مشابه برای $\tilde{\theta}(0/3)$ برابر است با $1/68$. همان‌طور که انتظار داریم، برآورد از طریق نمونه‌گیری مجموعه رتبه‌دار دقت بیشتری دارد. لازم به ذکر است که انتخاب $t = 0/3$ در بالا تنها برای نشان دادن کاربرد رویه پیشنهادی است. در عمل، انتخاب t بر مبنای اطلاعات جانبی موجود درباره مسئله مورد بررسی، مانند مطالعات پیشین یا نظر کارشناسان انجام می‌شود.



شکل (۴): برآورد میانگین مربع خطای $\hat{\theta}(t)$ و $\tilde{\theta}(t)$ تحت توزیع نمایی دومتغیره برای $n = 25$

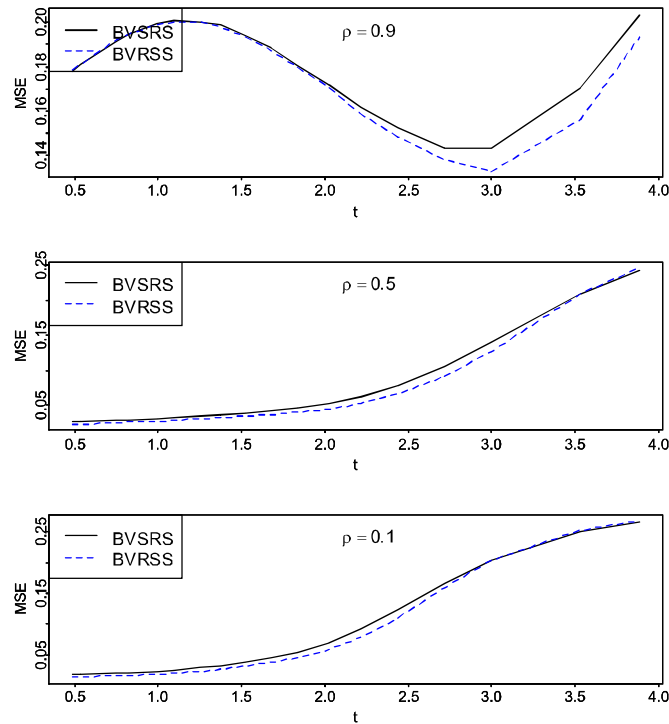


شکل (۵): برآورد میانگین مربع خطای $\hat{\theta}(t)$ و $\tilde{\theta}(t)$ تحت توزیع گامای دومتغیره برای $n = 9$.

۵- نتیجه‌گیری

استنباط آماری در مورد پارامتر $\theta = P(X > Y)$ موضوعی مهم در قابلیت اعتماد بوده که مورد توجه محققان زیادی قرار گرفته است. برآورد این پارامتر در زمینه‌های دیگری مانند پزشکی، روانشناسی و علوم تربیتی نیز کاربرد دارد. این مقاله به مسئله برآورد معیار قابلیت اعتماد پویای $\theta(t)$ بر اساس نمونه‌گیری تصادفی ساده دومتغیره و نمونه‌گیری مجموعه رتبه‌دار دومتغیره می‌پردازد. ویژگی‌های دو برآوردگر معرفی شده به صورت نظری و عددی بررسی شده‌اند، و کاربرد آن‌ها با استفاده از یک مجموعه داده پزشکی توضیح داده شده است. نتایج نشان می‌دهد برآورد حاصل از طرح نمونه‌گیری دوم غالباً میانگین مربع خطای کمتری دارد.

هر یک از برآوردگرهای پیشنهادی به صورت نسبت دو برآوردگر بر اساس تابع توزیع تجربی هستند. برآوردگرهای ساخته شده بر اساس تابع توزیع تجربی، دارای آریبی زیاد در دم توزیع هستند زیرا این گونه برآوردگرها نمی‌توانند در مقادیر t فراتر از بزرگ‌ترین مشاهده، برآورد دقیقی ارائه دهند (زاردشت و همکاران [۱۴]). به عنوان آینده تحقیق قصد داریم از برآورد هسته‌ای تابع چگالی در ساختن برآوردگرهای جدید برای $\theta(t)$ در دو طرح نمونه‌گیری فوق استفاده کنیم.



شکل (۶): برآورد میانگین مربع خطای $\hat{\theta}(t)$ و $\tilde{\theta}(t)$ تحت توزیع گامای دومتغیره برای $n = 25$

تشکر و قدردانی

نویسندگان از داوران ارجمند که نظرات سازنده‌شان موجب بهبود مقاله شد، و از سردبیر محترم مجله مدل‌سازی پیشرفته ریاضی صمیمانه سپاسگزارند.

منابع

- [1] McIntyre, G.A. (1952). A method of unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, **3**, 385-390.
- [2] Chen, Z., Bai, Z., and Sinha, B.K. (2004). *Ranked set sampling: theory and applications*. Springer, New York.
- [3] Al-Saleh, M.F., and Zheng, G. (2002). Estimation of bivariate characteristics using ranked set sampling. *Australian and New Zealand Journal of Statistics*, **44**, 221-232.
- [4] Al-Saleh, M.F., and Samawi, H.M. (2005). Estimation of the correlation coefficient using bivariate ranked set sampling with application to the bivariate normal distribution. *Communications in Statistics: Theory and Methods*, **34**, 875-889.
- [5] Kotz, S., Lumelskii, Y., and Pensky, M. (2003). *The stress-strength model and its generalizations. Theory and Applications*. World Scientific, Singapore.
- [6] Sengupta, S., and Mukhuti, S. (2008). Unbiased estimation of $P(X > Y)$ using ranked set sample data. *Statistics*, **42**, 223-230.
- [7] Samawi, H.M., Helu, A., Rochani, H.D., Yin, J., and Linder, D. (2016). Estimation of $P(X > Y)$ when X and Y are dependent random variables using different bivariate sampling schemes. *Communications for Statistical Applications and Methods*, **23**, 385-397.
- [8] Zardasht V., and Asadi, M. (2010). Evaluation of $P(X_t > Y_t)$ when both X_t and Y_t are residual lifetimes of two systems. *Statistica Neerlandica*, **64**, 460-481.
- [9] Mahdizadeh, M., and Zamanzade, E. (2016). A new reliability measure in ranked set sampling. *Statistical Papers* (DOI: 10.1007/s00362-016-0794-3).
- [10] Shao, J. (2003). *Mathematical statistics*, 2nd Edition. Springer, New York.
- [11] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8**, 229-231.
- [12] Walldius, G., Jungner, I., Aastveit, A.H., Holme, I., Furberg, C.D., and Sniderman, A.D. (2004). The apoB/apoA-I ratio is better than the cholesterol ratios to estimate the balance between plasma proatherogenic

- and antiatherogenic lipoproteins and to predict coronary risk. *Clinical Chemical Laboratory Medicine*, **42**, 1355-1363.
- [13] Walldius, G., and Jungner, I. (2006). The apoB/apoA-I ratio: a strong, new risk factor for cardiovascular disease and a target for lipid-lowering therapy: A review of the evidence. *Journal of Internal Medicine*, **259**, 493-519.
- [14] Zardasht, V., Zeepongsekul, P., and Asadi, M. (2012) On nonparametric estimation of a reliability function. *Communications in Statistics: Theory and Methods*, **41**, 983-999.

Estimation of Dynamic Reliability for Dependent Stress and Strength Based on Ranked Set Sampling

Mahdi Mahdizadeh* and Ehsan Zamanzade**

*Department of Statistics, Hakim Sabzevari University, Sabzevar, Iran

**Department of Statistics, University of Isfahan, Isfahan, Iran

Abstract

In this article, estimation of a dynamic reliability measure based on ranked set sampling is studied when stress and strength variables are dependent. The proposed estimator is compared with the analogous estimator in simple random sampling theoretically and numerically. The results show that estimation using ranked set sampling design is often more precise. Also, the difference in precision of the two estimators is more pronounced as the correlation coefficient between stress and strength variables goes higher.

Keywords: Judgment ranking, Ranked set sampling, Reliability.

Mathematics Subject Classification (2010): 62G30, 62N05