مجله علوم آماری، پاییز و زمستان ۱۳۸۹
جلد ۴، شماره ۲، ص ۲۳۱–۲۵۱

برآوردگر بهبود یافته ضریب تعیین در توزیع نرمال چند متغیره

احد ملک‌زاده، مینا توحیدی
گروه آمار، دانشگاه شیراز

تاریخ دریافت: ۱۳۸۸/۱۲/۱۰ **تاریخ آخرین بازنگری:** ۱۳۸۹/۸/۷

چکیده : مسئله برآورد نقطه‌ای ضریب تعیین در توزیع نرمال p متغیره مورد توجه افراد زیادی قرار گرفته است. این معیار به‌دلیل کاربرد فراوان، دارای اهمیت زیادی است، در این مقاله با در نظر گرفتن کلاس برآوردگرهای خطی دو برآوردگر جدید معرفی می‌شوند که دارای مخاطره کمتری نسبت به دو برآوردگر معمول یعنی ضریب تعیین نمونه‌ای و تعدیل شده آن باشند. همه برآوردگرهای ارائه شده اریب می‌باشند. بنابراین با معرفی برآوردگر جک‌نایف و مقایسه مخاطره این برآوردگر به‌وسیله شبیه‌سازی نشان داده می‌شود. برآوردگر جک‌نایف برآوردگری بهتری نسبت به برآوردگرهای دیگر می‌باشد.

واژه‌های کلیدی : ضریب تعیین، توزیع نرمال چند متغیره، برآورد جک‌نایف، زیان توان دوم خطا.

آدرس الکترونیک مسئول مقاله: احد ملک‌زاده، admalekzadeh@yahoo.com
کد موضوع‌بندی ریاضی (۲۰۱۰): ۶۲B۱۰

۱ مقدمه

فرض کنید $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ یک نمونه تصادفی n تائی از توزیع نرمال p متغیره با بردار میانگین $\boldsymbol{\mu}$ و ماتریس کواریانس $\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_{11} & \boldsymbol{\sigma}_1{}' \\ \boldsymbol{\sigma}_1 & \boldsymbol{\Sigma}_{22}\end{pmatrix}$ باشد. ضریب همبستگی چند گانه X_1 و بردار $\boldsymbol{X}_2=(X_2,\ldots,X_p)$ در حقیقت بیشترین همبستگی بین X_1 و بهترین ترکیب خطی از $\boldsymbol{X}_2$ می‌باشد و دارای اهمیت ویژه‌ای در علم روانشناسی، پزشکی، اقتصاد و علوم اجتماعی است.این معیار همبستگی برابراست با $\rho=\frac{\sqrt{\boldsymbol{\sigma}'_1\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_1}}{\sqrt{\sigma_{11}}}$ از آنجا که برآوردگرهای ماکسیمم درستنمایی $\boldsymbol{\mu}$ و $\boldsymbol{\Sigma}$ به‌صورت

$$\hat{\boldsymbol{\mu}}=\bar{X} \quad , \quad \hat{\boldsymbol{\Sigma}}=\frac{1}{n}A=\begin{pmatrix}a_{11} & \boldsymbol{a}'_{(1)} \\ \boldsymbol{a}_{(1)} & A_{22}\end{pmatrix}$$

هستند، که در آن $A=\sum_{\alpha=1}^{n}(\boldsymbol{X}_\alpha-\bar{\boldsymbol{X}})'(\boldsymbol{X}_\alpha-\bar{\boldsymbol{X}})$ است، برآورد ماکسیمم درستنمایی پارامتر ρ نیز برابر است با

$$R=\frac{\sqrt{\boldsymbol{a}_{(1)}{}'A_{22}^{-1}\boldsymbol{a}_{(1)}}}{\sqrt{a_{11}}}$$

اگرچه برآوردگر ماکسیمم درستنمائی R^2، به‌دلیل سادگی محاسبه کاربرد فراوانی دارد، اما این برآوردگر اریب بوده و همیشه مقدار ρ^2 را بیش از مقدار واقعی برآورد می‌کند. فیشر (۱۹۴۲) با روش گشتاوری و استفاده از تقریب مرتبه اول $E(R^2)$، یک برآوردگر تعدیل یافته به‌صورت

$$R_A^2=\frac{(n-1)R^2-(p-1)}{n-p},$$

معرفی نمود. پس از او افراد دیگری مانند ازکیل (۱۹۳۰) و وری (۱۹۳۱)، برآوردگر R_A^2 را به‌صورت‌های دیگر نیز معرفی کردند. مورهد (۱۹۸۵) مسئله برآورد $\theta=\rho^2/(1-\rho^2)$ را مورد توجه قرار داد و با در نظر گرفتن کلاس تمام برآوردگرهای خطی به‌صورت $b+a\frac{R^2}{1-R^2}$، با فرض $\mathrm{a},\mathrm{b}\in\Re$ ، نشان داد که این کلاس، برآوردگر نااریب با کمترین واریانس (UMVU) برای θ را نیز در بر دارد. مرشاند (۲۰۰۱) کلاس برآوردگرهای خطی از R^2 را مورد بررسی قرار داد که در

بخش ۲ به آن پرداخته خواهد شد. الکین و پرت (۱۹۵۸) نیز برآوردگر UMVU برای $\rho^۲$ را به‌صورت

$$U(r^۲) = ۱ - \frac{n-۳}{n-p}(۱-r^۲)\ {}_۲F_۱(۱,۱;\frac{n-p+۱}{۲}, ۱-r^۲),$$

ارائه کردند، که در آن تابع فوق هندسی ${}_pF_q(a;b;t)$ با دو بردار $a=(a_۱,a_۲,\ldots,a_p)$ و $b=(b_۱,\ldots,b_q)$ به‌صورت

$${}_pF_q(a;b;t) = \sum_{k=\circ}^{\infty} \frac{(a)_k t^k}{(b)_k k!}$$

است، که در آن

$$(a)_k = \prod_{i=۱}^{p}\prod_{j=\circ}^{k-۱}(a_i+j), \quad (b)_k = \prod_{i=۱}^{q}\prod_{j=\circ}^{k-۱}(b_i+j), \quad (a)_\circ = ۱, \quad (b)_\circ = ۱$$

تعریف می‌شود (رینوایل، ۱۹۶۷). اگرچه برآوردگرهای $R^۲$ و $R_A^۲$ به‌دلیل سادگی محاسبه، کاربرد فراوانی دارند، مرشاند (۲۰۰۱) ثابت کرد تحت زیان توان دوم خطا $R_A^۲$ در هر بعدی ناپذیرفتنی است و $R^۲$ در ابعاد کمتر از ۷ در کلاس برآوردگرهای پذیرفتنی قرار می‌گیرد. دربخش ۲ به‌منظور یافتن برآوردگرهای بهینه ($R_{Mb}^۲$ و $R_{Mr}^۲$)، به بررسی کلاس برآوردگرهای خطی ارائه شده توسط مرشاند (۲۰۰۱) پرداخته می‌شود. در بخش ۳ و برآوردگر جک‌نایف $R^۲$ معرفی و تعدیل خواهد شد. در بخش ۴ نتایج شبیه‌سازی برای به‌دست آوردن مخاطره برآوردگرهای جک‌نایف تحت تابع زیان توان دوم خطا ارائه خواهد شد و در انتها به بحث و نتیجه‌گیری پرداخته می‌شود. همچنین نمودارهایی به منظور درک مخاطره‌های جدید ارائه گردیده است.

۲ کلاس برآوردگرهای خطی $\rho^۲$

فرض کنید $\boldsymbol{X}_۱,\ldots,\boldsymbol{X}_n$ نمونه‌ای تصادفی از توزیع نرمال p متغیره با بردار میانگین نامعلوم $\boldsymbol{\mu}$ و ماتریس کوواریانس نامعلوم $\sum$ باشد. $R^۲$ ضریب تعیین نمونه‌ای به‌دست آمده از این نمونه در نظر گرفته می‌شود. در ادامه کلاس برآوردگرهای

خطی از $R^۲$ به‌صورت $D = \left\{\delta_{a,b} \middle| \delta_{a,b}(R^۲) = aR^۲ + b \quad a, b \in \Re\right\}$ را مورد بررسی قرار داده و عملکرد این برآوردگرها تحت تابع زیان توان دوم خطا $L(\rho^۲, \delta_{a,b}) = (\delta_{a,b} - \rho^۲)^۲$ که در آن $۱ \leq \rho^۲ \leq ۰$ ، مورد ارزیابی قرار می‌گیرد، یعنی با فرض $a = ۱ - b = ۱$ یا $a = ۱ - b = \frac{n-۱}{n-p}$ به‌ترتیب دو برآوردگر $R^۲$ و $R_A^۲$ به‌دست خواهند آمد. این کلاس اولین بار توسط مرشاند (۲۰۰۱) مورد بررسی قرار گرفت که در ادامه بخشی از نتایج کار وی ذکر خواهد شد.

توزیع $R^۲$ برای $\rho^۲ = ۱$ توزیعی تباهیده در نقطه ۱ است. درنتیجه مخاطره این برآوردگر نیز در نقطه $\rho^۲ = ۱$ برابر صفر می‌شود. بنابراین باید برآوردگرهای خطی از $R^۲$ را در نظر گرفت که مقدار مخاطره آن‌ها در نقطه یک برابر صفر باشد، که نتیجه می‌دهد $a + b = ۱$. بنابراین کافی است زیر کلاسی از D به‌صورت $D_۱ = \left\{\delta_a \middle| \delta_a(R^۲) = aR^۲ + ۱ - a, \quad a \in \Re\right\}$ مورد بررسی قرار گیرد. تابع مخاطره برآوردگرهای موجود در کلاس $D_۱$ ، توسط مرشاند (۲۰۰۱) مورد ارزیابی قرار گرفت و نتایج زیر به‌دست آمد. (در تمامی حالات باید $n > p$ باشد.)

لم ۱ (مرشاند، ۲۰۰۱): تابع مخاطره هر برآوردگر $\delta_a \in D_۱$ به‌صورت

$$R(\rho^۲, \delta_a) = (a - \alpha(\rho^۲))^۲ E(Z^۲) + (۱ - \rho^۲)^۲ - \alpha(\rho^۲) E(Z^۲),$$

است، که در آن $Z = ۱ - R^۲$ و

$$\alpha(\rho^۲) = \frac{(n+۱)(۱-\rho^۲)^۲ {}_۲F_۱(۱, ۱; \frac{n+۱}{۲}; \rho^۲)}{(n-p+۲)\ {}_۲F_۱(۲, ۲; \frac{n+۳}{۲}; \rho^۲)}$$

تابعی نزولی روی بازه (۱ و ۰) می باشد، به‌طوری که برای $n \geq ۶$

$$\bar{\alpha} = \sup\{\alpha(\rho^۲)\ ;\ \rho^۲ \in [۰, ۱)\} = \frac{n+۱}{n-p+۲}$$

$$\underline{\alpha} = \inf\{\alpha(\rho^۲)\ ;\ \rho^۲ \in [۰, ۱)\} = \frac{n-۵}{n-p+۲}$$

قضیه ۱ (مرشاند، ۲۰۰۱):

الف) برای $n \geq ۶$ ، زیرکلاس مینیمال برای کلاس $D_۱$ به‌صورت

$$D_۱^* = \left\{ \delta_a \in D_۱ \left| \frac{n-۵}{n-p+۲} \leq a \leq \frac{n+۱}{n-p+۲} \right. \right\}$$

می‌باشد. ب) برای $n \geq ۶$، $R_A^۲$ برآوردگری ناپذیرفتنی است و در زیرکلاس مینیمال $D_۱^*$ ، توسط δ_aهائی با

$$\max\{\frac{(n+۱)(n-p)-۲(p-۱)}{(n-p)(n-p+۵)}, \frac{n-۵}{n-p+۲} \leq a \leq \frac{n+۱}{n-p+۲}\},$$

مغلوب خواهد شد.

ج) برای $n > p \geq ۸$ برآوردگر $R^۲$ ناپذیرفتنی است و در زیرکلاس مینیمال $D_۱^*$ توسط δ_a با

$$\frac{n-۵}{n-p+۲} \leq a \leq \min\{\frac{n+۱}{n-p+۲}, \frac{n+p-۱۲}{n-p+۲}\},$$

مغلوب خواهد شد.

در اینجا میزان اریبی برآوردگرهای موجود در کلاس $D_۱$ بررسی می‌شود. با توجه به تابع چگالی و گشتاورهای مختلف $R^۲$ که توسط مورهد (۱۹۸۲) ارائه شده، داریم

$$E(\delta_a) = ۱ - a\frac{n-p}{n-۱}\,(۱-\rho^۲)\ \ {}_۲F_۱(۱, ۱; \frac{n+۱}{۲}; \rho^۲)$$

و می‌توان لم زیر را درباره میزان اریبی برآوردگرهای موجود در کلاس $D_۱$ بیان کرد.

لم ۲ : اگر مقدار اریبی هر برآوردگر $\delta_a \in D_۱$ را با $b_a(n\ ,\mathrm{p}\ ,\rho^2)$ نمایش داده شود، آنگاه

الف) برای هر (n, p) ، اگر $a \leq \frac{n-۳}{n-p}$ آنگاه $b_a(n\ ,\mathrm{p}\ ,\rho^2)$ تابعی محدب، نزولی از ρ^2 بوده و δ_a یک بیش برآوردگر می‌باشد.

ب) برای هر (n, p) ، اگر $a > \frac{n-۱}{n-p}$ آنگاه $b_a(n\ ,\mathrm{p}\ ,\rho^2)$ تابعی محدب، صعودی از ρ^2 می‌باشد، که مقادیر منفی می‌پذیرد.

با توجه به لم ۲ و اینکه به ازاء هر بردار ثابت (n, p, ρ^2)، تابع $b_a(n, \mathrm{p}, \rho^2)$ در نقطه $\rho^۲ = ۱$ مقدار صفر را می‌پذیرد، پس با انتخاب یک مقدار a در فاصله $\left[\frac{n-۳}{n-p}, \frac{n-۱}{n-p}\right]$، می‌توان به برآوردگری با قدر مطلق اریبی کمتر دست یافت. با پیشنهاد نقطه وسط این فاصله، یعنی $a_۱ = \frac{n-۲}{n-p}$، برآوردگر جدید به‌صورت

$$R^۲_{Mb} = \frac{n-۲}{n-p}R^۲ - \frac{p-۲}{n-p}$$

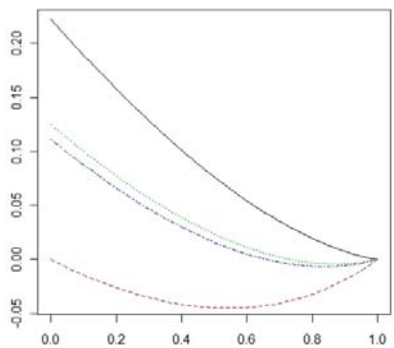
معرفی می‌شود، که به ازاء $p = ۲$ همان $R^۲$ خواهد بود.

بنابر قسمت (الف) قضیه ۱، برای $n \geq ٦$، بازهٔ مناسب برای a، در کلاس مینیمال $D^*_۱$، $[\frac{n-۵}{n-p+۲}, \frac{n+۱}{n-p+۲}]$ است. اگر $n > ۲p + ۳$ باشد، اشتراک این بازه با بازهٔ به‌دست آمده برای a به‌منظور یافتن برآوردگری با قدر مطلق اریبی کم، بازه $[\frac{n-۳}{n-p}, \frac{n+۱}{n-p+۲})$می‌شود. به‌عنوان مثال، می‌توان یک نقطه $a_۰ = \frac{n-۱}{n-p+۱}$ را درون این بازه انتخاب کرد و برآوردگر دیگری به‌صورت

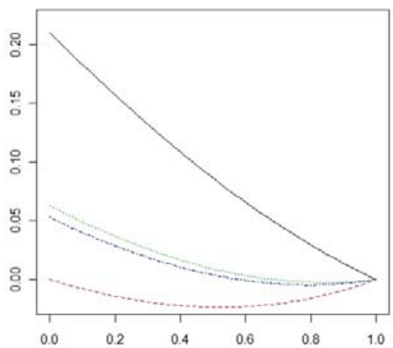
$$R^۲_{Mr} = \frac{n-۱}{n-p+۱}R^۲ - \frac{p-۲}{n-p+۱}$$

معرفی نمود، که به ازاء $p = ۲$، همان برآوردگر $R^۲$ است.

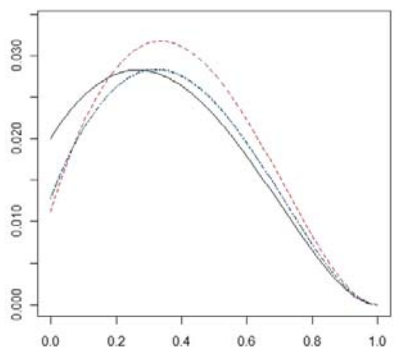
شکل ۱ اریبی و مخاطره برآوردگرهای $R^۲$، $R^۲_A$، $R^۲_{Mr}$ و $R^۲_{Mb}$ را در چند حالت مختلف نشان می‌دهد. شکل ۱، نشان می‌دهد که به ازاء هر n و p برآوردگر $R^۲_A$ دارای اریبی منفی است و در دو نقطه صفر و یک مقدار اریبی آن نیز صفر می‌شود البته این مطلب با توجه به قسمت (ب) لم ۲ نیز قابل درک است. برآوردگر $R^۲$ نیز به ازاء $p > ۲$ همیشه بیش برآوردگر می‌باشد. دو برآوردگر دیگر یعنی $R^۲_{Mr}$ و $R^۲_{Mb}$ اریبی معتدل‌تری نسب به دو برآوردگر $R^۲$ و $R^۲_A$ دارند. این دو برآوردگر به غیر نقطه ۱ در نقطه‌ای بزرگتر از صفر نیز مقدار اریبی صفر می‌پذیرند. شکل‌های ۲ تا ۵ مخاطره این سه برآوردگر را نمایش می‌دهند. دو برآوردگر جدید $R^۲_{Mr}$ و $R^۲_{Mb}$ در همه نقاط مخاطره ای کمتر از برآوردگر $R^۲_A$ دارند این موضوع برای $p > ۷$ نیز در مورد $R^۲$ صدق می‌کند. در بعدهای کمتر مساوی ۷ مخاطره این سه برآوردگر قابل مقایسه نیستند. در ضمن با افزایش بعد، مخاطره برآوردگر $R^۲$ نیز افزایش می‌یابد.



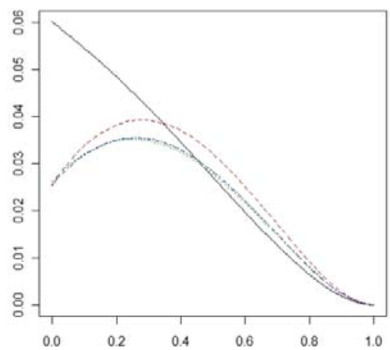
(الف)



(ب)



(ج)



(د)

شکل ۱: ، الف: اریبی $p = ۴, n = ۱۰$، ب: اریبی $n = ۲۰, p = ۵$، ج: مخاطره $n = ۲۰, p = ۳$ و د: مخاطره $n = ۲۰, p = ۵$

۳ برآورد جک‌نایف ضریب تعیین نمونه‌ای

روش جک‌نایف از جمله روش‌های باز نمونه‌گیری و ناپارامتری است و به عنوان یک ابزار قوی برای برآورد واریانس یک برآوردگر نااریب یا برآورد واریانس و مقدار اریبی یک برآوردگر اریب مورد استفاده قرار می‌گیرد. در این روش براساس زیر نمونه‌هایی از نمونهٔ اصلی برآوردهایی از پارامترهای جامعه را می‌توان نتیجه گرفت (برای اطلاعات بیشتر به مقالات افرون (۱۹۷۹ و ۱۹۹۱) مراجعه شود).
فرض کنید $\hat{\theta}$ برآوردگر اریب به دست آمده از یک نمونه n تائی برای پارامتر θ باشد. برای کاهش اریبی این برآوردگر $\hat{\theta}_i$ را مقدار به‌دست آمده از فرمول $\hat{\theta}_i$ برای تمام نمونه بجز نمونه iام در نظر گرفته و تعریف می‌شود $\bar{\theta} = \sum\limits_{i=1}^{n} \hat{\theta}_i/n$. آنگاه برآوردگر مقدار اریبی به روش جک‌نایف به‌صورت $(n-1)(\bar{\theta}-\hat{\theta})$ به‌دست می‌آید. اگر برآورد جک‌نایف از مقدار $(n-1)(\bar{\theta}-\hat{\theta})$ کم شود برآوردگر تقریبا نااریبی به‌صورت

$$\hat{\theta}_{jack} = n\hat{\theta} - (n-1)\,\bar{\theta}$$

به‌دست می‌آید، که برآوردگر جک‌نایف نامیده می‌شود. برآورد خطای استاندارد به روش جک‌نایف به‌صورت

$$\hat{Se}_{\ jack} = \left[\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \bar{\theta})^2\right]^{1/2}$$

تعریف می‌شود. حال فرض کنید $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ یک نمونه تصادفی nتائی از توزیع $N_p(\mu, \Sigma)$ باشند. برای به‌دست آوردن برآوردگر جک‌نایف ρ^2 در کلاس D_1، δ_{ai} را مقدار برآوردگر δ_a به‌دست آمده از همهٔ نمونه‌ها به جز نمونه iام در نظر گرفته و $\bar{\delta}_a$ به‌صورت

$$\bar{\delta_a} = \frac{1}{n}\sum_{i=1}^{n}\delta_{ai}$$

معرفی می‌شود.

برآورد جک‌نایف مقدار اریبی برآوردگر δ_a متعلق به کلاس D_1 با توجه به رابطه $T_a = (n-1)(\bar{\delta_a} - \delta_a)$ تعریف می‌شود و برآورد جک‌نایف به‌صورت

$\delta_{ajack} = n\delta_a - (n-۱)\,\bar{\delta_a}$ است، که به ازاء $a = ۱$ و $a = \frac{n-۱}{n-p}$ برآوردهای جک‌نایف $R^۲$ و $R^۲_A$ به‌ترتیب به‌صورت

$$\begin{aligned} R^۲_{jack} &= nR^۲ - (n-۱)\,\bar{R^۲} \\ R^۲_{Ajack} &= nR^۲_A - (n-۱)\,\bar{R^۲_A} \end{aligned}$$

به‌دست می‌آیند. برآورد خطای استاندارد δ_{ajack} با نماد $\hat{Se}(\delta_{ajack})$ نشان داده می‌شود و و برابر است با

$$\begin{aligned} S\hat{e}(\delta_{ajack}) &= \left[\frac{n-۱}{n}\sum_{i=۱}^{n}(\delta_{ai} - \bar{\delta_a})^۲\right]^{۱/۲} \\ &= \left[a^۲\frac{n-۱}{n}\sum_{i=۱}^{n}(R^۲_i - \bar{R^۲})^۲\right]^{۱/۲} \\ &= a\,S\hat{e}(R^۲_{jack}) \end{aligned}$$

قضیه ۲ : برآوردگر $R^۲_{Ajack}$ ناریب از مرتبه دو و برآوردگر $R^۲_{jack}$ ناریب از مرتبه اول هستند، به‌عبارت دیگر

$$E(R^۲_{Ajack}) = \rho^۲ + o(n^{-۲}) \quad E(R^۲_{jack}) = \rho^۲ + o(n^{-۱})$$

برهان : با استفاده از امید ریاضی ضریب همبستگی نمونه‌ای و قرار دادن $x = \rho^۲$، داریم

$$\begin{aligned} E(R^۲_{Ajack}) &= nE(R^۲_A) - (n-۱)E(\bar{R}^۲_A) \\ &= n(۱ - (۱-x)\,{}_۲F_۱(۱,۱;\frac{n+۱}{۲};x)) \\ &- (n-۱)(۱ - (۱-x)\,{}_۲F_۱(۱,۱;\frac{n}{۲};x)) \\ &= ۱ + (۱-x)(n-۱ + \frac{۲x(n-۱)}{n} + \frac{۸x^۲(n-۱)}{n(n+۲)} \end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{48x^3(n-1)}{n(n+2)(n+4)+\ldots)} \\
&- \quad (n+\frac{2xn}{n+1}+\frac{8x^2n}{(n+1)(n+3)}+\frac{48x^3n}{(n+1)(n+3)(n+5)+\ldots})
\end{aligned}
$$

پس از ساده کردن داریم

$$E(R^2_{Ajack}) = x + \frac{2x(1-x)(4x-1)}{n^2} + o(n^{-2}) = x + o(n^{-2})$$

به‌طور مشابه

$$
\begin{aligned}
E(R^2_{jack}) &= nE(R^2) - (n-1)E(R^2) \\
&= x + \frac{3(1-x)}{n} + \frac{(1-x)(8x^2 - 2px + 2x - (p+1))}{n^2} + o(n^{-2}) \\
&= x + \frac{3(1-x)}{n} + O(n^{-2}) \\
&= x + o(n^{-1})
\end{aligned}
$$

با توجه به قضیه ۲، اگر در تعریف R^2_{jack} ضرائب n و $n-1$ به‌ترتیب با $n-1$ و $n-2$ عوض شوند برآوردگر جدید

$$R^2_{jacknew} = (n-1)R^2 - (n-2)\bar{R}^2$$

نیز ناریب از مرتبه دو می‌شود. بنابر شبیه‌سازی‌های انجام شده، ملاحظه می‌شود که برآوردگرهای معرفی شــده در این بخش برای ρ^2 مقادیر منفی را نیز اختیار می‌کنند. برای رفع این مشکل، تعریف می‌شود $\delta^+_{ajack} = \max(0, \delta_{ajack})$ و به همین صورت دو برآوردگر R^{2+}_{jack} و $R^{2+}_{jacknew}$ نیز معرفی می‌شوند. از آن‌جا که توزیع R^2 در نقطه $\rho^2 = 1$ تباهیده می‌باشد تمام برآوردگرهای معرفی شده در این بخش به ازای $\rho^2 = 1$ مقداری برابر با ۱ خواهند داشت.

۴ مطالعه شبیه‌سازی

به‌دلیل پیچیدگی توزیع δ_{ajack}، برای مقایسه مخاطره برآوردگرها، از شبیه‌سازی استفاده شده است. به‌منظور بررسی کارائی برآوردهای جک‌نایف، ۱۰۰۰۰ نمونه از

توزیع نرمال p متغیره در حالت‌های مختلف شبیه‌سازی و از نتایج حاصل برای مخاطره برآوردگرهای مختلف، یافتن نقاط بحرانی برای آزمون فرضیه $H_\circ : \rho^۲ = \circ$ و برآورد واریانس $R^۲$ استفاده می‌شود.

به‌منظور بررسی رفتار برآوردگرهای جک‌نایف ارائه شده، شبیه‌سازی سنگین و وقت‌گیری را متحمل شدیم که نتایج را در سه زمینه مقایسه مخاطره این برآوردگرها با سایر برآوردگرها ($R^۲$، $R_A^۲$، $R_{Mr}^۲$ و $R_{Mb}^۲$)، یافتن نقاط بحرانی برای آزمون $H_\circ : \rho^۲ = \circ$ و محاسبه ضریبی به‌منظور برآورد واریانس $R^۲$ توسط برآوردگرهای $R_i^۲$ به‌کار گرفته می‌شود. در استفاده از برآوردگرهای جک‌نایف مختلف تنها از برآوردگرهای جک‌نایف $R_{jack}^{۲+}$ و $R_{jacknew}^{۲+}$ استفاده خواهد شد، زیرا شبیه‌سازیهای انجام‌شده، نشان می‌دهد این دو برآوردگر تحت تابع زیان درجه دوم عملکرد بهتری نسبت به مابقی دارند. بنابراین برای مقایسه برآوردگرهای $R^۲$، $R_A^۲$، $R_{Mr}^۲$ و $R_{Mb}^۲$ با برآوردگرهای جک‌نایف، تنها این برآوردگرها با $R_{jack}^{۲+}$ و $R_{jacknew}^{۲+}$ مقایسه خواهند شد. در شکل ۳ مخاطره برآوردگرهای جک‌نایف در مقابل چهار برآوردگر دیگر در حالت‌های مختلف رسم شده است. درتمامی حالات (به ازاء ابعاد مختلف نرمال چند متغیره) و تعداد نمونه کم برآوردگرهای جک‌نایف برآوردگرهائی با عملکرد بهتر هستند و با زیاد شدن تعداد نمونه کارائی آن‌ها نسبتا کمتر می‌شود. می‌توان گفت وقتی تعداد نمونـه زیاد باشد، بهتر است از $R_{Mr}^۲$ استفاده شود. در جدول ۲ نیز مقدار مخاطره این برآوردگرها در بعدهای مختلف و اندازه‌های نمونه متفاوت آورده شده است.

منظور در ستون سوم جدول ۱ (e) به‌جای $R^۲$، $R_A^۲$، $R_{Mr}^۲$، $R_{Mb}^۲$، $R_{jack}^{۲+}$ و $R_{jacknew}^{۲+}$ از اعداد ۱ تا ۶ درج شده‌اند. جدول ۱ نشان‌دهنده این موضوع است که در بعد ۳، برآوردگر $R^۲$ مخاطره کمتری را داراست در حالی که با افزایش بعد ($p > ۷$) مخاطره این برآوردگر بطور قابل ملاحظه‌آی بیشتر از مخاطره مابقی برآوردگرها می‌شود. در تمام نقاط، تمام ابعاد و هر حجم نمونه ای $R_A^۲$ مخاطره‌ای بیشتر از دو برآوردگر $R_{Mr}^۲$ و $R_{Mb}^۲$ دارد. در حجم نمونه‌های کم (p بزرگ و n کوچک) برآوردگر $R_{jack}^{۲+}$ دارای مخاطره‌ای کمتر از مابقی برآوردگرهاست.

جدول ۱: مخاطره برآوردگرهای مختلف در برآورد ضریب همبستگی جامعه

۰	۰/۳	۰/۱۱	۰/۴	۰/۶۵	۰/۷۵	۰/۹۴	e	n	p
۰/۰۸۰۸	۰/۰۷۳۹	۰/۰۶۰۷	۰/۰۵۲۶	۰/۰۳۳۴	۰/۰۱۷۴	۰/۰۰۱۵	۱		
۰/۰۵۱۹	۰/۰۶۵۴	۰/۰۷۴۴	۰/۰۷۱۸	۰/۰۵۲۳	۰/۰۲۸۹	۰/۰۰۲۵	۲		
۰/۰۵۳۴	۰/۰۵۸۸	۰/۰۶۰۰	۰/۰۵۶۲	۰/۰۳۹۸	۰/۰۲۱۸	۰/۰۰۱۹	۳	۱۰	
۰/۰۵۵۴	۰/۰۵۹۵	۰/۰۵۹۲	۰/۰۵۵۱	۰/۰۳۸۶	۰/۰۲۱۱	۰/۰۰۱۸	۴		
۰/۰۴۷۷	۰/۰۵۶۹	۰/۰۷۰۲	۰/۰۷۵۲	۰/۰۵۹۶	۰/۰۳۵۱	۰/۰۰۳۳	۵		
۰/۰۴۹۲	۰/۰۵۷۷	۰/۰۶۹۲	۰/۰۷۳۰	۰/۰۵۷۰	۰/۰۳۵۵	۰/۰۰۳۱	۶		
۰/۰۰۵۰	۰/۰۱۰۵	۰/۰۱۴۲	۰/۰۱۳۷	۰/۰۰۹۳	۰/۰۰۴۷	۰/۰۰۰۴	۱		۳
۰/۰۰۲۶	۰/۰۰۹۸	۰/۰۱۵۱	۰/۰۱۴۹	۰/۰۱۰۴	۰/۰۰۵۳	۰/۰۰۰۴	۲		
۰/۰۰۳۲	۰/۰۰۹۶	۰/۰۱۴۳	۰/۰۱۴۰	۰/۰۰۹۷	۰/۰۰۴۹	۰/۰۰۰۴	۳	۴۰	
۰/۰۰۳۲	۰/۰۰۹۶	۰/۰۱۴۲	۰/۰۱۴۰	۰/۰۰۹۷	۰/۰۰۴۹	۰/۰۰۰۴	۴		
۰/۰۰۲۲	۰/۰۱۰۵	۰/۰۱۶۳	۰/۰۱۵۸	۰/۰۱۰۴	۰/۰۰۵۵	۰/۰۰۰۴	۵		
۰/۰۰۲۳	۰/۰۱۰۵	۰/۰۱۶۲	۰/۰۱۵۸	۰/۰۱۰۴	۰/۰۰۵۵	۰/۰۰۰۴	۶		
۰/۲۴۲۴	۰/۱۹۴۱	۰/۱۲۴۴	۰/۰۹۴۲	۰/۰۴۵۸	۰/۰۱۹۷	۰/۰۰۱۳	۱		
۰/۱۴۵۵	۰/۱۴۴۱	۰/۱۲۹۰	۰/۱۱۴۶	۰/۰۷۴۳	۰/۰۳۸۶	۰/۰۰۳۲	۲		
۰/۱۲۷۳	۰/۱۲۰۹	۰/۱۰۳۱	۰/۰۹۰۰	۰/۰۵۷۲	۰/۰۲۹۵	۰/۰۰۲۴	۳	۱۰	
۰/۱۲۸۸	۰/۱۱۸۲	۰/۰۹۶۳	۰/۰۸۲۶	۰/۰۵۱۱	۰/۰۲۶۰	۰/۰۰۲۱	۴		
۰/۰۹۰۷	۰/۰۸۵۵	۰/۰۸۷۱	۰/۰۸۷۳	۰/۰۶۹۹	۰/۰۴۱۹	۰/۰۰۳۸	۵		
۰/۰۹۶۸	۰/۰۸۹۹	۰/۰۸۷۵	۰/۰۸۶۰	۰/۰۶۵۶	۰/۰۳۸۶	۰/۰۰۳۵	۶		۵
۰/۰۱۵۰	۰/۰۱۷۴	۰/۰۱۷۲	۰/۰۱۵۴	۰/۰۰۹۶	۰/۰۰۴۶	۰/۰۰۰۳	۱		
۰/۰۰۵۶	۰/۰۱۲۱	۰/۰۱۶۶	۰/۰۱۶۰	۰/۰۱۰۹	۰/۰۰۵۵	۰/۰۰۰۴	۲		
۰/۰۰۶۰	۰/۰۱۱۸	۰/۰۱۵۷	۰/۰۱۵۱	۰/۰۱۰۲	۰/۰۰۵۱	۰/۰۰۰۴	۳	۴۰	
۰/۰۰۶۰	۰/۰۱۱۹	۰/۰۱۵۷	۰/۱۵۰	۰/۰۱۰۲	۰/۰۰۵۱	۰/۰۰۰۴	۴		
۰/۰۰۳۹	۰/۰۱۱۰	۰/۰۱۷۷	۱/۰۱۷۵	۰/۰۱۱۰	۰/۰۰۵۷	۰/۰۰۰۴	۵		
۰/۰۰۴۰	۰/۰۱۱۱	۰/۰۱۷۶	۰/۰۱۷۴	۰/۰۱۱۰	۰/۰۰۵۷	۰/۰۰۰۴	۶		
۰/۴۴۲۰	۰/۳۴۸۷	۰/۲۱۴۶	۰/۱۵۷۴	۰/۰۷۰۰	۰/۰۲۵۷	۰/۰۰۱۶	۱		
۰/۲۲۵۰	۰/۲۰۳۱	۰/۱۵۷۷	۰/۱۳۰۸	۰/۰۷۴۴	۰/۰۳۵۲	۰/۰۰۲۶	۲		
۰/۰/۱۹۹۱	۰/۱۷۷۹	۰/۱۳۶۲	۰/۱۱۲۳	۰/۰۶۳۴	۰/۰۲۹۸	۰/۰۰۲۲	۳	۱۵	
۰/۱۸۴۰	۰/۱۶۰۳	۰/۱۱۸۳	۰/۰۹۶۰	۰/۰۵۲۷	۰/۰۲۴۵	۰/۰۰۱۸	۴		
۰/۰۹۷۲	۰/۰۸۵۴	۰/۰۸۱۷	۰/۰۷۹۷	۰/۰۶۰۵	۰/۰۳۳۴	۰/۰۰۲۵	۵		
۰/۱۰۵۳	۰/۰۹۱۵	۰/۰۸۴۰	۰/۰۷۸۹	۰/۰۵۷۰	۰/۰۳۰۶	۰/۰۰۲۷	۶		
۰/۰۳۹۶	۰/۰۳۴۶	۰/۰۲۵۲	۰/۰۲۰۲	۰/۰۱۰۵	۰/۰۰۴۶	۰/۰۰۰۳	۱		۱۰
۰/۰۰۸۸	۰/۰۱۳۴	۰/۰۱۵۶	۰/۰۱۴۵	۰/۰۰۹۴	۰/۰۰۴۶	۰/۰۰۰۳	۲		
۰/۰۰۸۹	۰/۰۱۳۰	۰/۰۱۵۰	۰/۰۱۳۹	۰/۰۰۸۹	۰/۰۰۴۴	۰/۰۰۰۳	۳	۵۰	
۰/۰۰۹۰	۰/۰۱۳۰	۰/۰۱۴۹	۰/۰۱۳۸	۰/۰۰۸۹	۰/۰۰۴۴	۰/۰۰۰۳	۴		
۰/۱۰۵۳	۰/۰۹۱۵	۰/۰۸۴۰	۰/۰۷۸۹	۰/۰۵۷۰	۰/۰۳۰۶	۰/۰۰۲۵	۵		
۰/۰۰۵۸	۰/۰۱۱۹	۰/۰۱۶۰	۰/۰۱۵۰	۰/۰۰۹۶	۰/۰۰۴۷	۰/۰۰۰۳	۶		

جدول ۲: مخاطره برآوردهای مختلف در برآورد واریانس ضریب همبستگی نمونه‌ای R^2

p	n	e	۰/۹۴	۰/۷۵	۰/۶۵	۰/۴	۰/۱۱	۰/۳	۰
		۱	۱/۰۵۴	۶/۵۰۶	۲۰/۲۱۳	۲۲/۱۲۷	۱۴/۹۵۰	۴/۸۳۱	۸/۰۲۹
		۲	۲/۱۸۸	۱۱/۳۳۸	۱۶/۲۵۴	۲۳/۶۰۶	۳۵/۷۱۷	۴۳/۴۶۱	۳۷/۹۸۶
	۱۰	۳	۱/۵۵۷	۸/۷۲۵	۱۷/۴۷۳	۱۸/۸۹۴	۱۷/۴۲۷	۱۱/۹۰۹	۹/۱۱۳
		۴	۱/۴۸۷	۸/۴۲۴	۱۷/۷۴۷	۱۸/۹۶۶	۱۶/۴۷۶	۱۰/۰۹۸	۸/۰۳۸
		۵	۱/۴۸۰	۸/۷۳۵	۲۰/۶۸۱	۲۲/۰۶۲	۱۷/۴۹۷	۷/۰۹۵	۴/۱۳۲
		۶	۳/۷۵۳	۲۳/۶۵۹	۷۵/۷۴۲	۹۴/۴۲۴	۱۰۴/۸۲۱	۱۱۶/۲۵۵	۱۳۸/۱۲
۳		۱	۰/۰۰۶	۰/۱۱۰	۰/۶۵۹	۰/۳۷۱	۰/۳۷۱	۰/۹۴۱	۱/۵۵۴
		۲	۰/۰۰۸	۰/۱۳۵	۰/۶۳۵	۰/۳۲۹	۰/۸۰۹	۱/۹۱۲	۱/۰۶۵
	۴۰	۳	۰/۰۰۷	۰/۱۲۱	۰/۶۴۲	۰/۳۳۹	۰/۵۴۷	۱/۲۷۹	۱/۰۲۲
		۴	۰/۰۰۷	۰/۱۲۱	۰/۶۴۲	۰/۳۴۰	۰/۵۴۲	۱/۲۶۶	۱/۰۰۲۹
		۵	۰/۰۰۶	۰/۱۱۶	۰/۶۶۹	۰/۳۷۳	۰/۷۴۷	۱/۵۱۰	۰/۶۶۲
		۶	۰/۰۰۸	۰/۱۷۵	۱/۸۴۷	۲/۱۸۱	۲/۴۰۴	۲/۸۹۴	۱/۹۸۰
		۱	۰/۲۳۶	۲/۳۵۹	۱۹/۱۲۶	۳۶/۷۶۶	۳۹/۹۶۳	۳۵/۷۸۰	۲۷/۹۶۳
		۲	۱/۴۵۴	۹/۱۲۴	۱۸/۶۹۹	۱۷/۴۶۴	۱۷/۴۶۴	۲۶/۷۳۲	۲۰/۷۹۰
	۱۰	۳	۱/۰۱۸	۶/۸۸۳	۱۷/۸۳۹	۱۸/۸۳۳	۱۸/۴۹۶	۱۶/۶۰۱	۱۴/۸۴۲
		۴	۰/۸۳۴	۵/۸۸۱	۱۷/۵۵۲	۲۰/۵۷۰	۲۰/۳۵۳	۱۶/۶۲۱	۱۳/۱۸۵
		۵	۰/۹۷۱	۶/۶۹۸	۲۰/۵۷۸	۲۳/۶۰۱	۲۲/۷۹۸	۱۷/۲۴۹	۱۲/۶۰۶
۵		۶	۰/۹۷۱	۶/۶۸۹	۲۰/۵۷۸	۲۳/۶۰۱	۲۲/۷۹۸	۱۷/۲۴۹	۱۲/۶۰۶
		۱	۰/۰۰۴	۰/۰۸۴	۰/۶۸۸	۰/۴۹۳	۰/۲۸۳	۰/۶۲۴	۲/۴۴۸
		۲	۰/۰۰۷	۰/۱۲۴	۰/۶۸۸	۰/۳۴۴	۰/۵۱۵	۱/۷۸۸	۱/۶۳۸
	۴۰	۳	۰/۰۰۶	۰/۱۱۱	۰/۶۸۸	۰/۳۵۱	۰/۳۹۲	۱/۱۸۳	۱/۲۵۷
		۴	۰/۰۰۶	۰/۱۱۰	۰/۶۸۸	۰/۳۵۲	۰/۳۸۴	۱/۱۴۵	۱/۲۴۹
		۵	۰/۰۰۶	۰/۱۰۷	۰/۶۵۴	۱/۳۸۷	۰/۴۷۲	۰/۰۴۱	۰/۷۰۰
		۶	۰/۰۰۷	۰/۱۴۹	۱/۷۲۰	۲/۴۶۰	۲/۲۷۸	۲/۹۲۳	۳/۳۵۷
		۱	۰/۰۰۹	۰/۱۹۴	۵/۱۲۸	۱۴/۷۳۳	۲۲/۹۷۳	۳۳/۲۷۵	۳۹/۳۳۳
		۲	۰/۱۵۰	۱/۴۹۱	۸/۵۲۶	۱۰/۱۴۹	۱۰/۴۹۱	۱۱/۱۱۲	۱۲/۸۷۶
	۱۵	۳	۰/۱۱۷	۱/۱۷۹	۷/۳۱۹	۹/۵۴۱	۹/۷۵۷	۹/۴۳۳	۹/۵۰۴۱
		۴	۰/۰۸۰	۰/۸۳۴	۵/۹۳۱	۸/۹۲۰	۹/۸۹۷	۱۰/۲۷۱	۱۰/۳۰۸
		۵	۰/۰۹۷	۱/۰۱۱	۶/۳۸۱	۸/۴۴۱	۸/۸۲۴	۹/۱۴۶	۹/۵۴۵
		۶	۰/۰۵۷	۰/۰۷۷۳	۱۲/۹۲۰	۳۲/۳۲۶	۴۳/۱۲۶	۶۳/۹۷۲	۸۲/۰۵۸
۱۰		۱	۰/۰۰۱	۰/۰۳۱	۰/۴۵۹	۰/۵۶۳	۰/۲۸۰	۰/۱۲۰	۱/۴۴۰
		۲	۰/۰۰۳	۰/۰۴۸	۰/۳۴۹	۰/۲۰۰	۰/۳۲۲	۰/۷۱۴	۱/۰۸۸
	۵۰	۳	۰/۰۰۲	۰/۰۴۴	۰/۳۴۳	۰/۲۰۹	۰/۱۶۸	۰/۵۰۲	۰/۸۱۱
		۴	۰/۰۰۲	۰/۰۴۴	۰/۳۴۲	۰/۲۱۴	۰/۱۶۳	۰/۴۶۸	۰/۷۷۴
		۵	۰/۰۰۲	۰/۰۰۴۷	۰/۳۵۳	۰/۲۱۶	۰/۱۷۳	۰/۳۱۱	۰/۲۹۴
		۶	۰/۰۰۲	۰/۰۵۱	۰/۷۵۸	۱/۲۰۴	۱/۳۵۲	۱/۹۶۶	۳/۱۸۴

دومین استفاده از برآوردگر جک‌نایف آزمون صفر بودن ضریب تعیین جامعه در سطوح مختلف می‌باشد. بدین منظور در جدول ۲ تعدادی از نقاط بحرانی مربوطه ذکر شده است. به‌دلیل اینکه رفتار R^{2+}_{jack} همانند رفتار نیست می‌توان با استفاده از روش بونفرونی آزمون $H_0 : \rho^2 = 0$ را در سطح α به‌صورت زیر انجام داد.

براساس یک نمونه n تائی از توزیع نرمال p متغیره اگر هر دو برآوردگر R^2 و R^{2+}_{jack} در سطح $\alpha/2$ فرض صفر را تأیید کردند آنگاه فرضیه $H_0 : \rho^2 = 0$ پذیرفته می‌شود. در غیر این‌صورت فرض صفر رد می‌شود.

ستون سوم جدول ۲، احتمال درست برآورد کردن ضریب تعیین جامعه هنگامی که مقدار آن صفر می‌باشد را نمایش می‌دهد. همان‌طور که ملاحظه می‌شود با افزایش تعداد نمونه، احتمال برآوردیابی درست R^{2+}_{jack} آن نیز بیشتر می‌شود.

سومین کاربرد برآوردگر جک‌نایف یافتن ضریبی مناسب برای برآورد واریانس برآوردگر R^2 یا هر برآوردگر خطی از آن می‌باشد. بدین منظور براساس یک نمونه nتائی از توزیع نرمال p متغیره برآورد $Var(R^2)$ به‌صورت

$$Var_{jack} = \frac{k}{n-1}\sum_{i=1}^{n}(R^2_i - \bar{R^2})^2$$

پیشنهاد می‌شود: هدف، یافتن k بگونه‌ای است که Var_{jack} برآوردگری تقریباً نااریب برای واریانس R^2 باشد. بر اساس شبیه‌سازی‌های انجام شده می‌توان نتیجه گرفت، که برای تعداد نمونه بیش از ۲۵، $R = 1, n-p-4$ و در بقیه حالات $k \approx 0/85n - 0/7p - 1/5$ است.

نتایج شبیه‌سازی بیانگر آن است با افزایش نمونه، واریانس معرفی‌شده برآوردگری نااریب برای تمام بازه $\rho^2 \in (0,1)$ می‌باشد، ولی برای تعداد نمونه کم برای بازه $\rho^2 \in (0.4,1)$ برآوردگری نااریب و برای مابقی نقاط یک بیش برآوردگر است.

واریانس برآوردگر R^2 دارای فرم بسته است. به‌عنوان برآوردگرهای دیگری از واریانس R^2، اگر مقادیر پنج برآوردگر R^2، R^2_A، R^2_{Mr}، R^2_{Mb}، R^{2+}_{jack} بجای ρ^2 در فرمول مربوطبه واریانس R^2 قرار داده شود، آنگاه پنج برآوردگر $\hat{Var}_{R^2}$، $\hat{Var}_{R^2_A}$، $\hat{Var}_{R^2_{Mb}}$، $\hat{Var}_{R^2_{Mr}}$ و $\hat{Var}_{jack}$ به‌دست خواهد آمد. برای بررسی عملکرد برآوردگر

Var_{jack} با این برآوردگرها، مخاطره هر یک براساس ۱۰۰۰۰ بار شبیه‌سازی محاسبه و نتایج در جدول ۳ ارائه و در شکل ۲ نشان داده شده است.

اعداد ۱ تا ۶ در ستون سوم جدول ۳ (e) به‌ترتیب نشان‌دهندهٔ $\hat{Var}_{R^2}$، $\hat{Var}_{R_A^2}$، $\hat{Var}_{R_{Mb}^2}$، $\hat{Var}_{R_{Mr}^2}$، $\hat{Var}_{jack}$ و Var_{jack} می‌باشد.

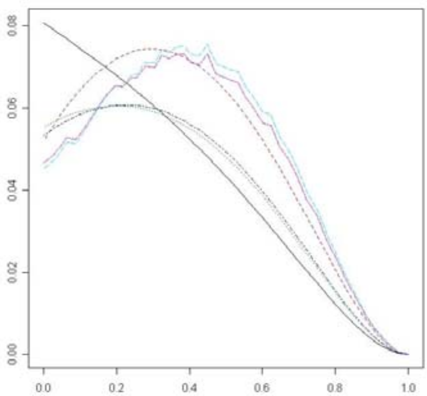
جدول ۳: نقاط بحرانی R_{jack}^{2+} برای آزمون فرض $H_0 = \rho^2 = 0$ و احتمال درست برآورد کردن ضریب تعیین جامعه تحت صفر

پیش‌گویی			α			
p	n	$\rho^2 = 0$	۰/۱	۰/۰۵	۰/۰۲۵	۰/۰۱
۳	۱۰	۰/۵۶۹۱	۰/۳۹۳۵	۰/۵۲۶۰	۰/۶۲۴۷	۰/۷۳۵۸
	۲۰	۰/۵۹۲۶	۰/۱۶۱۳	۰/۲۳۱۵	۰/۳۰۲۱	۰/۳۷۰۸
	۴۰	۰/۶۱۶۴	۰/۰۷۶۲	۰/۱۱۴۳	۰/۱۴۶۳	۰/۱۸۸۲
	۵۰	۰/۶۱۶۷	۰/۰۵۹۳	۰/۰۸۷۸	۰/۱۱۹۶	۰/۱۵۴۱
۵	۱۰	۰/۴۹۳۶	۰/۵۶۸۴	۰/۶۸۴۸	۰/۷۲۲۸	۰/۸۵۹۳
	۲۰	۰/۵۵۴۷	۰/۲۴۲۳	۰/۳۲۵۶	۰/۳۹۸۹	۰/۴۸۲۳
	۴۰	۰/۵۶۶۷	۰/۱۰۹۸	۰/۱۵۵۱	۰/۱۶۶۷	۰/۲۵۲۷
	۵۰	۰/۵۷۵۵	۰/۰۸۵۶	۰/۱۱۹۹	۰/۱۴۹۰	۰/۱۹۰۳
۱۰	۱۰	۰/۴۵۶۹	۰/۵۹۳۰	۰/۶۹۴۰	۰/۷۷۳۴	۰/۸۴۰۳
	۲۰	۰/۴۸۳۹	۰/۴۰۴۸	۰/۵۰۳۴	۰/۵۸۷۹	۰/۶۷۵۵
	۴۰	۰/۵۳۳۴	۰/۱۷۱۲	۰/۲۲۷۱	۰/۲۷۶۰	۰/۳۳۴۶
	۵۰	۰/۵۳۴۷	۰/۱۳۵۲	۰/۱۸۰۳	۰/۲۲۳۳	۰/۲۷۳۷

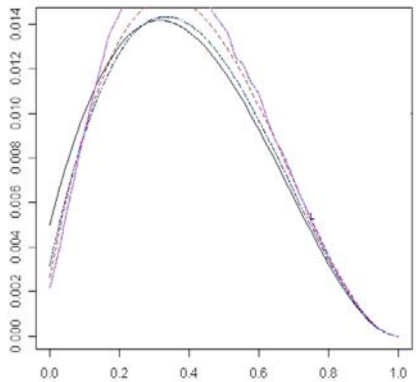
بحث و نتیجه‌گیری

از بحث‌های ارائه شده در بخش‌های قبل، به‌طور کلی می‌توان نتایج زیر را به‌دست آورد: برآوردگر R^2 در $p = 2$ برآوردگری بهینه است و در بعدهای بالاتر و تعداد نمونه کم $(p < n < 2p)$ بهترین برآوردگر دو برآوردگر جک‌نایف می‌باشند. به منظور سادگی محاسبات، در حالتی که $p \geq 5$ می‌توان از برآوردگر R_{Mr}^2 برای ρ^2استفاده کرد.

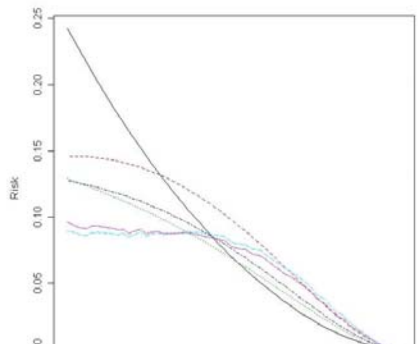
همچنین از برآوردگر جک‌نایف R_{jack}^2 براساس جدول ۳ می‌توان برای آزمون



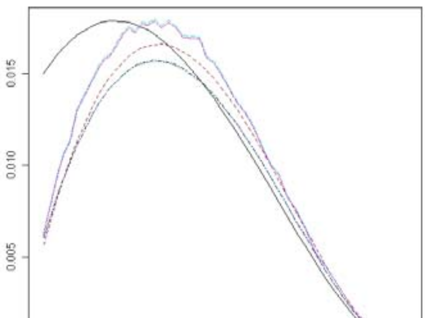
(الف)



(ب)

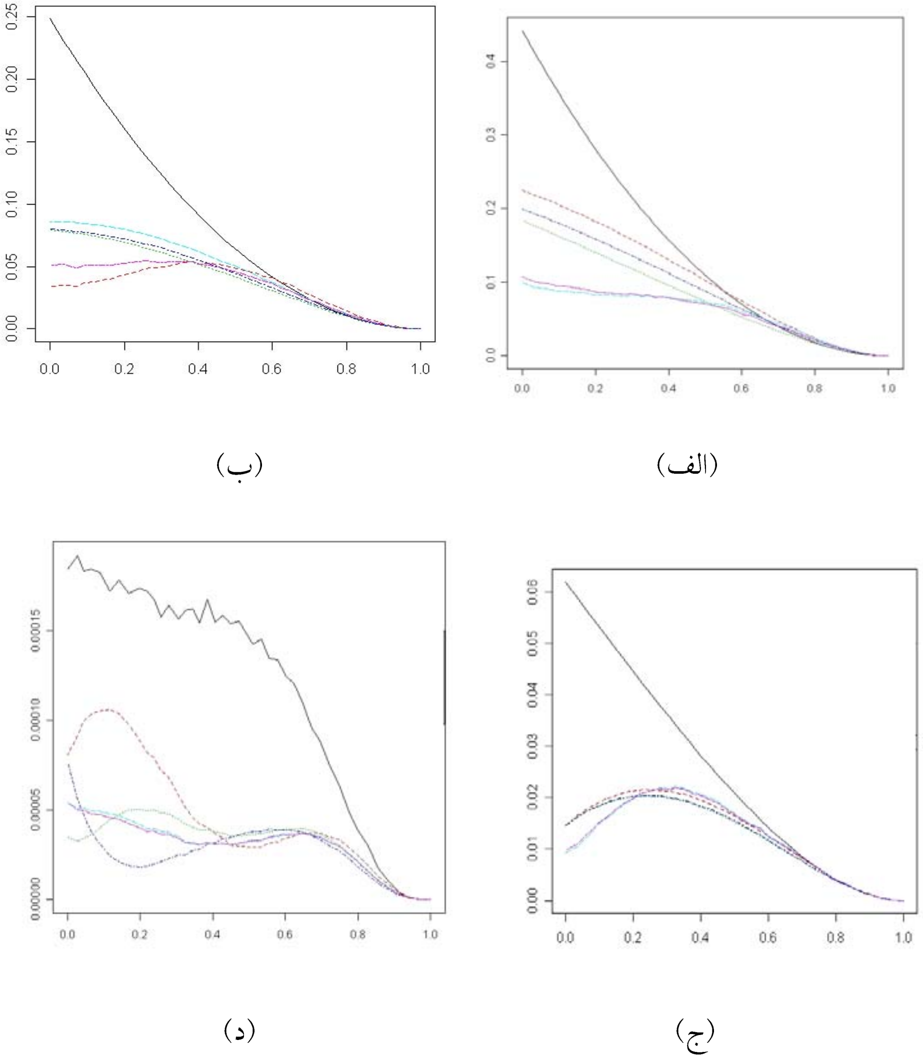


(ج)

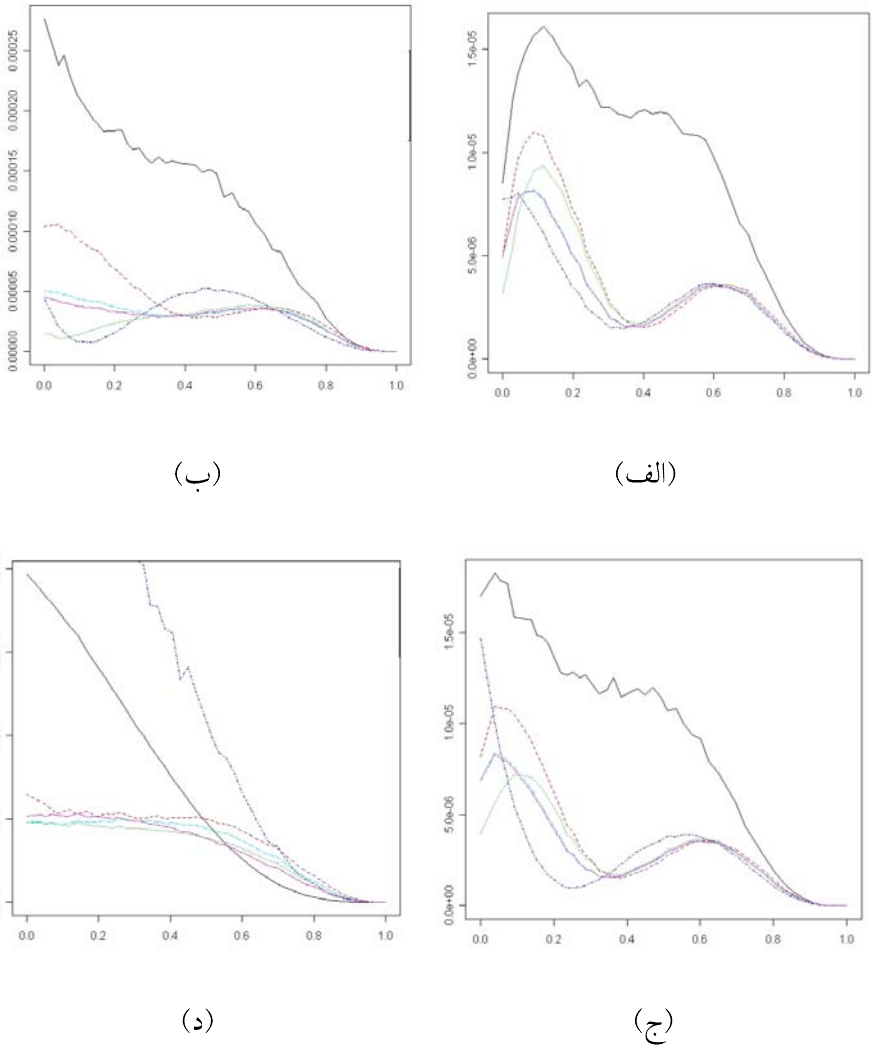


(د)

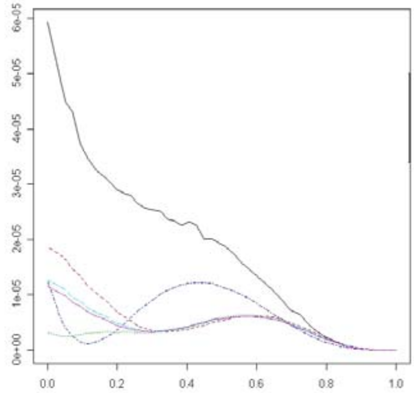
شکل ۲: ، الف: مخاطره $p = ۳, n = ۱۰$، ب: $p = ۳, n = ۴۰$، ج: $p = ۵, n = ۱۰$ و د: $p = ۵, n = ۴۰$



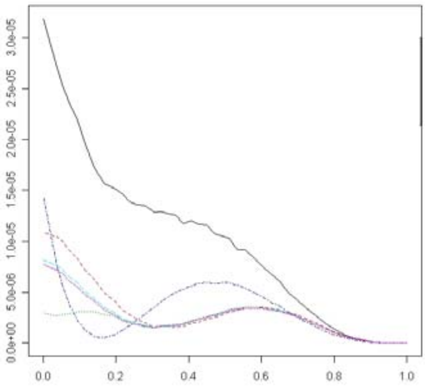
شکل ۳: الف: مخاطره $p = ۱۰, n = ۱۵$، ب: $n = ۲۰, p = ۱۰$، ج: $n = ۴۰, p = ۱۰$ و د: $n = ۲۰, p = ۳$



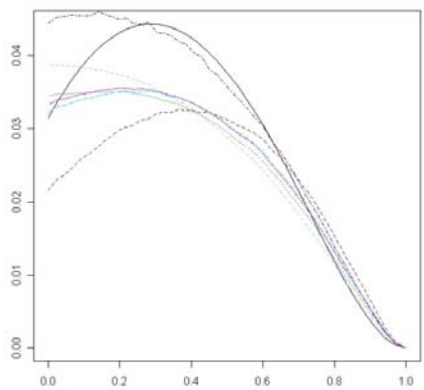
شکل ۴: الف: مخاطره $p = ۳, n = ۵۰$، ب: $n = ۲۰, p = ۵$ ج: $n = ۵۰, p = ۵$ و د: $n = ۱۵, p = ۱۰$



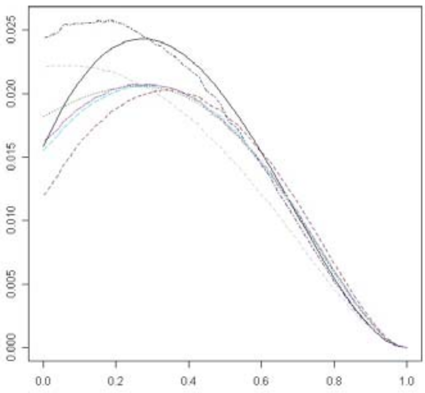
(الف)



(ب)



(ج)



(د)

شکل ۵: ، الف: مخاطره $p = ۱۰, n = ۴۰$، ب: $n = ۵۰, p = ۱۰$، ج: $n = ۱۰, p = ۳$ و د: $n = ۲۰, p = ۵$

فرضیه صفر بودن ضریب تعیین جامعه، در همه حالات مطابق با آن‌ها در بخش ۳ ذکر شد استفاده کرد.

به‌عنوان برآورد مناسبی برای واریانس R^2 نیز می‌توان از $\hat{Var}_{jack}$ استفاده نمود، و استفاده از برآورد معرفی‌شده در بخش ۳، برآورد توصیه‌شده با بهره‌گیری از روش جک‌نایف، مناسب نیست.

تشکر و قدردانی

نویسندگان از پیشنهادهای ارزنده داوران محترم کمال تشکر را دارند.

مراجع

Efron, B., (1979), Bootsrap Methods: Another Look at the Jacknife. *Annals of Statistics*, **7**, 1-26.

Efron, B., (1981), Nonparametric Estimator of Standard Error: The Jackknife, the Bootstrap and other Methods, *Biometrika*,**68**, 589-599.

Ezekiel, M., (1930), *Methods of Correlational Analysis*, Wiley, New York.

Fisher, R. A., (1924), The Influence of Rainfall on the Yield of Wheat at Rehashed, *Philosophical Transactions of Royal Society of London*, Series B, **213**, 89-124.

Fisher, R. A., (1928), The General Sampling Distribution of the Multiple Correlation Coefficients, *Proc Royal Society*, Series A, **121**, 654-673.

Marchand, E., (2001), Point Estimation of the Coefficient of Determination, *Statistics & Decisions*, **19**, 137-154.

Muirhead, R. J., (1982), *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Muirhead, R. J., (1985), Estimating a Particular Function of the Multiple Correlation Ccoefficients, *Journal of the American Statistical Association*, **80**, 923-925.

Olkin, I. and Pratt, J. W, (1958), Unbiased Estimation of Certain Correlation Coefficient, *Annals of Mathematical Statistics*, **29**, 201-211.

Rainville, E. D., (1967), *Special Functions, The Macmillan Company*, New York.

Wherry, R. J. Sr., (1931), A New Formula for Predicting the Shrinkage of Multiple Correlations, *Annals of Mathematical Statistics*, **2**, 440-457.

269 Journal of Statistical Sciences, Autumn & Winter 2010
Vol. 4, No. 2, pp 231-251.

Improved Estimator of Coefficient of Determination in Multivariate Normal Distribution

Malekzadeh, A. and Tohidi, M.

Statistics Deptartment, Shiraz University, Shiraz, Iran.

Abstract: Coefficient of determination is an important criterion in different applications. The problem of point estimation of this parameter has been considered by many researchers. In this paper, the class of linear estimators of $R^{٢}$ was considered. Then, two new estimators were proopposed, which have lower risks than other usual estimators, such as the sample coefficient of determination and its adjusted form. Also on the basis of some simulations, we show that the Jackknife estimator is an efficient estimator with lower risk, when the number of observations is small.

Keywords: Coefficient of determination, Multivariate Normal Distribution, Jackknife estimator, Squared error loss function.

Mathematics Subject Classification (2000): 62B10