

تشخیص و حذف هم خطی در مدل رگرسیونی با استفاده از نظریه اطلاع

فائزه شکیبا^۱، غلامرضا محتشمی برزادران^۲

تاریخ دریافت: ۱۳۹۳/۱۲/۱۷

تاریخ پذیرش: ۱۳۹۵/۷/۱۷

چکیده:

در این مقاله نخست مسئله هم خطی در مدل رگرسیونی معرفی می شود و شیوه تشخیص هم خطی و راه های برطرف کردن هم خطی مطرح می شوند. در ادامه تعاریف مقدماتی از نظریه اطلاع عنوان می شوند که در نهایت با استفاده از نظریه اطلاع، هم خطی در مدل رگرسیونی شناسایی می شود و راه حلی برای برطرف کردن آن پیشنهاد می شود. **واژه های کلیدی:** رگرسیون، هم خطی، نظریه اطلاع، توزیع پسین.

۱ مقدمه

ذکر است که اگر هیچ رابطه ای خطی بین متغیرهای پیشگو وجود نداشته باشد، متغیرهای پیشگو متعامد هستند و پیش بینی و انتخاب یک مجموعه مناسب از متغیرها برای مدل به آسانی صورت می پذیرد. یکی از راه های تشخیص هم خطی، ماتریس همبستگی است (مونتگومری [۱۲]) که می توان برای برطرف کردن هم خطی از مدل رگرسیونی ستیغی استفاده کرد که توسط هورل [۶] معرفی شده است. در رگرسیون ستیغی، ضرایب رگرسیونی از رابطه

$$\beta_R = (X'X + \lambda I)^{-1} X'Y, \quad (1)$$

به دست می آیند. در رابطه (۱)، $\lambda \geq 0$ است که توسط تحلیل گر انتخاب می شود. برآورد ضرایب مدل رگرسیونی ستیغی، برآوردی اریب با کمترین واریانس است که ضرایب رگرسیونی را برآورد می کند.

با اختراع تلفن، مفهوم اطلاع و انتقال، افزایش یافت که برخی از دانشوران به بررسی مشکلات انتقال اطلاع پرداختند. تحقیق در مورد نظریه اطلاع، اولین بار توسط هری نیکوئست [۱۳] در مقاله ای به نام «عوامل خاصی که سرعت تلگراف را تحت تأثیر قرار می دهند» صورت گرفت. کار مهم دیگر در این زمان، انتشار مقاله «انتقال اطلاع» توسط هارتلی [۵] بود که اولین پایه های ریاضی نظریه اطلاع را بنا گذاشت. تولد واقعی نظریه اطلاع را می توان به مقاله «نظریه ریاضی مخابرات» نوشته کلود شانون

واژه رگرسیون از لحاظ لغوی در فرهنگ لغت به معنی پس روی، برگشت و بازگشت است؛ اما از دید آمار و ریاضیات به مفهوم بازگشت به یک مقدار متوسط یا میانگین به کار می رود. مسائلی از نوع رگرسیون در قرن هجدهم مورد توجه قرار گرفت. از رگرسیون برای به الگو در آوردن رابطه بین متغیرهای آماری استفاده می شود. می توان رگرسیون را این گونه معرفی کرد که روابط نادقیق بین متغیرهای آماری را تعیین می کند و به تحلیل این روابط می پردازد. در رگرسیون، متغیر پاسخ و پیشگو داریم (متغیرهای پاسخ و پیشگو به ترتیب به عنوان متغیرهای وابسته و مستقل شناخته می شوند). در این مقاله متغیر پاسخ را پیوسته در نظر گرفته ایم؛ بنا بر این مدل های خطی رگرسیون را به کار می بریم. اگر در مدل رگرسیونی فقط یک متغیر پیشگو باشد مدل رگرسیونی، مدل رگرسیونی خطی ساده است که برابر است با $Y = \beta_0 + \beta_1 X + \varepsilon$. β_0 عرض از مبدأ، β_1 شیب خط و ε_i ها خطاها می باشند که دارای توزیع نرمال و همچنین مستقل هستند.

حال اگر در مدل رگرسیونی k متغیر پیشگو داشته باشیم، با مدل رگرسیونی خطی چندگانه سروکار داریم که به صورت $Y = X\beta + \varepsilon$ است. یک مسئله جدی که می تواند استفاده از مدل رگرسیونی را با اشکال مواجه کند، مسئله هم خطی یا همبستگی خطی نزدیک بین متغیرهای پیشگو است. قابل

^۱ دانش آموخته کارشناسی ارشد آمار، دانشگاه فردوسی مشهد

^۲ استاد گروه آمار، دانشگاه فردوسی مشهد

می‌توان ثابت کرد که $K(f : g) \geq 0$ است و تساوی برقرار خواهد بود اگر و فقط اگر $f(x) = g(x)$ باشد. همچنین قابل توجه است که نابرابری $K(f : g) \neq K(g : f)$ برقرار است؛ یعنی آنتروپی کولبک-لیبلر خاصیت تقارن ندارد (کاور و توماس [۳]). حاصل جمع $K(f : g)$ و $K(g : f)$ اطلاع جفریز [۸] را نتیجه می‌دهد که از رابطه

$$J(f : g) = K(f : g) + K(g : f)$$

حاصل می‌شود.

علاوه بر این، آنتروپی کولبک-لیبلر بر حسب آنتروپی شانون به صورت

$$K(f : g) = -H(X) - E_f(\log g(X)) \quad (۳)$$

خواهد بود. به عنوان مثال، اگر f و g دو تابع چگالی نرمال، به ترتیب با بردارهای میانگین و ماتریس واریانس-کوواریانس عبارت است از: (μ_f, Σ_f) و (μ_g, Σ_g) باشند، اطلاع کولبک-لیبلر برابر است با

$$K(f : g) = \frac{1}{4}(\text{tr}(\Sigma_f \Sigma_g^{-1}) - n) - \frac{1}{4} \log \frac{|\Sigma_f|}{|\Sigma_g|} + \frac{1}{4}(\mu_f - \mu_g)' \Sigma_g^{-1}(\mu_f - \mu_g) \quad (۴)$$

که در آن tr اثر ماتریس می‌باشد.

۳ آنتروپی ماکسیم

کاوش و توسعه بحث آنتروپی از سال ۱۹۴۸ تا کنون تحولات بسیاری را موجب شده است که بحث آنتروپی ماکسیم یکی از آنهاست. آنتروپی ماکسیم که توسط جینز [۷] مطرح شد، اشاره به این مطلب دارد که در انتخاب توزیع برای متغیر، باید فقط و فقط از اطلاعاتی استفاده کنیم که داریم. به این معنی که باید توزیعی را در نظر بگیریم که تحت اطلاع داده شده بیشترین عدم قطعیت یا کمترین تصادفی بودن را داشته باشد؛ چرا که اگر از توزیعی با عدم قطعیت کمتر استفاده کنیم به این معنی است که از اطلاعاتی استفاده کرده ایم که به ما داده نشده است، برای محاسبه آنتروپی ماکسیم شانون برای متغیر تصادفی X لازم است در ابتدا قضیه ۱.۳ مطرح شود که در این قضیه می‌توان توابع f و f' (مشتق تابع چگالی) را به دست آورد.

[۱۷] نسبت داد. مفهوم اطلاعی که توسط شانون مطرح شد از دید آمار و احتمال است و با مفاهیم روزمره از اطلاع، مانند دانش، تفاوت دارد.

حال می‌خواهیم هم‌خطی را با کمک نظریه اطلاع برطرف کنیم. تفسیر هم‌خطی بیزی توسط لیمر [۱۴] ارائه شده است. در ساختار بیزی علاقه‌مند به درک مجموعه‌ای از داده‌هایی هستیم که هم‌خطی دارند، و می‌خواهیم بدانیم که هم‌خطی چه اثرهایی بر توزیع پسین می‌گذارد.

۲ آنتروپی و اندازه‌های اطلاع

به طور کلی نظریه اطلاع، مدلی ریاضی از شرایط و عوامل مؤثر بر انتقال و پردازش داده‌ها و اطلاع فراهم می‌آورد. آنتروپی اطلاع که با نام آنتروپی شانون هم شناخته می‌شود، در واقع میزان تصادفی بودن یک رخداد را به صورت یک نسخه ریاضی گزارش می‌کند. بنا بر این برای متغیر تصادفی پیوسته X با تابع چگالی f ، آنتروپی از رابطه

$$H(X) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx$$

به دست می‌آید. قابل ذکر است که اگر X و Y متغیرهای تصادفی پیوسته با تابع چگالی توأم f باشند، آنتروپی توأم و آنتروپی شرطی از فرمول‌های

$$H(X, Y) = - E_f(\log f(X, Y))$$

$$H(X|Y) = - E_f(\log f(X|Y = y))$$

حاصل می‌شوند.

اگر X و Y از هم مستقل باشند، آن‌گاه

$$H(X, Y) = H(X) + H(Y)$$

و اگر X و Y هم‌توزیع باشند، تساوی $H(X) = H(Y)$ برقرار است. کولبک-لیبلر [۱۱] اندازه اطلاع کولبک-لیبلر را معرفی کردند که این اندازه را آنتروپی نسبی یا اطلاع واگرا هم می‌نامند. چون که این اندازه اختلاف بین دو توزیع از متغیر تصادفی را به دست می‌دهد، دارای اهمیت است. فرض کنید f و g دو تابع چگالی باشند. آنتروپی نسبی f و g به صورت زیر تعریف می‌شود:

$$K(f : g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx. \quad (۲)$$

۴ تشخیص هم خطی در مدل رگرسیونی خطی با استفاده از نظریه اطلاع

حال می‌خواهیم با کمک تعریف آنتروپی و اطلاع کولبک-لیبلر، هم خطی در مدل رگرسیونی را تشخیص دهیم. برای این منظور باید توزیع پسین ضرایب را داشته باشیم. اگر در مدل رگرسیونی با مسئله هم خطی مواجه باشیم، مدل رگرسیونی خطی را دوباره به صورت زیر بازنویسی می‌کنیم:

$$Y = (X\Gamma)(\Gamma'\beta) + \varepsilon = W\alpha + \varepsilon, \quad (5)$$

که $\Gamma = [\gamma_1, \dots, \gamma_p]$ ماتریس متعامد از ویژه بردارهای ماتریس $X'X$ و $W'W$ ماتریس قطری از ویژه مقدارهای (λ_j) ماتریس $X'X$ است. از این جا به بعد، ماتریس $W'W$ را با نماد Λ نشان می‌دهیم. با فرض این که داده‌ها و σ^2 (واریانس خطاها) معلوم باشند، توزیع پسین پارامتر β عبارت است از

$$\Pi_X(\beta|y, \sigma^2) = N(b, \sigma^2(X'X)^{-1}).$$

بنا بر این اندازه عدم اطمینان یا همان آنتروپی β به وسیله توزیع پسین از رابطه

$$H_X(\beta|y, \sigma^2) = \frac{p}{2} \log(2\pi e \sigma^2) - \log |X'X|^{\frac{1}{2}} \quad (6)$$

به دست می‌آید. مقادیر بزرگ $H(\beta|y)$ نشان دهنده این است که داده‌ها شامل اندازه کوچکی از اطلاع پارامتر β هستند. بنا بر این اثر هم خطی ضرایب رگرسیونی بر مبنای

$$I_X(\beta|\sigma^2) = -H_X(\beta|y, \sigma^2)$$

تعیین می‌شود. در نتیجه

$$I_X(\beta|\sigma^2) = \sum_{j=1}^p \log \lambda_j^{\frac{1}{2}} - \frac{p}{2} \log(2\pi e). \quad (7)$$

اگر X° نشان دهنده ماتریس متعامد و

$$\Pi_{X^\circ}(\beta|y, \sigma^2) = N(b^\circ, \sigma^2 I)$$

باشد، اثر هم خطی ضرایب رگرسیونی عبارت است از

$$I_{X^\circ}(\beta|\sigma^2) = \frac{p}{2} \log(2\pi e) + \frac{p}{2} \log(\sigma^2). \quad (8)$$

قضیه ۱.۳. اگر فرض کنیم رابطه

$$I = \int_a^b F(x, f(x), f'(x)) dx$$

را داشته باشیم (که در آن F یک تابع معلوم است)، به کمک معادله لاگرانژ اویلر، تابع $f(x)$ ای که I را ماکسیم می‌کند از تساوی زیر به دست می‌آید (وحیدیان و بزرگ‌نیا [۱]):

$$\frac{\partial F}{\partial f(x)} - \frac{d}{dx} \left(\frac{\partial F}{\partial f'(x)} \right) = 0.$$

به کمک قضیه ۱.۳، کاگان و همکاران [۹] قضیه زیر را معرفی کردند که با کمک آن می‌توان آنتروپی ماکسیم شانون را برای متغیر تصادفی X به دست آورد.

قضیه ۲.۳. هرگاه X یک متغیر تصادفی گسسته باشد که به ازای هر x متعلق به تکیه گاه D ، $p(x) > 0$ و همچنین توابع $r_i(x)$ برای $i = 1, \dots, k$ طوری باشند که $\sum_{x \in D} p(x) r_i(x) = \alpha_i$ و $\sum_{x \in D} p(x) = 1$ و α_i ها ثابت باشند، توزیع دارای آنتروپی ماکسیم به صورت نمایی

$$p^*(x) = e^{\lambda_0 + \lambda_1 r_1(x) + \dots + \lambda_k r_k(x)}$$

است، که $\lambda_0, \lambda_1, \dots, \lambda_k$ به کمک شرایط ذکر شده به دست می‌آیند (کاگان و همکاران [۹]).

چون در آنتروپی ماکسیم، تابع f' را نداریم، تابع زیر علامت انتگرال تابعی از x و $f(x)$ است و فقط $\frac{\partial F}{\partial f(x)} = 0$ باید حل شود.

مثال ۳.۳. آنتروپی ماکسیم n -متغیره X بر حسب محدودیت‌های $E(X) = \mu$ و $E(X - \mu)^2 = \Sigma$ به صورت زیر است:

$$f^*(x) = \left(\frac{1}{\sqrt{2\pi}} \right)^n |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}.$$

همچنین آنتروپی $N(\mu, \Sigma)$ برابر با

$$H(X) = \frac{n}{2} \log 2\pi e + \frac{1}{2} \log |\Sigma|$$

خواهد بود، که $|\Sigma|$ دترمینان ماتریس کوواریانس است و برآورد درست‌نمایی Σ با استفاده از چگالی آنتروپی ماکسیم عبارت است از

$$\hat{\Sigma} = \frac{1}{n} (X - \hat{\mu})'(X - \hat{\mu}).$$

قابل ذکر است که $\hat{\mu}$ برآورد درست‌نمایی μ در چگالی آنتروپی ماکسیم است.

رگرسیونی به صورت زیر به دست می آید:

$$H(\beta|\tau^2) = \frac{p}{2} \log(\pi e) + \frac{p}{2} \log \tau^2 \quad (11)$$

و همچنین فرض می کنیم که توزیع پسین $(\Pi(\beta|y, \sigma^2, m, \tau^2))$ ضرایب رگرسیونی برابر با $N(b(\phi, m), \sigma^2(X'X + \phi I_p)^{-1})$ باشد، که

$$b(\phi, m) = (X'X + \phi I_p)^{-1}(X'Y + \phi m), \quad (12)$$

و $\phi = \frac{\sigma^2}{\tau^2}$ نسبت دقت مدل است. اگر در رابطه (۱۲)، $m = 0$ باشد، ضرایب رگرسیونی مدل ستیغی به دست خواهد آمد. در نتیجه

$$H_X(\beta|y, \sigma^2, \tau^2) = \frac{p}{2} \log(\pi e) + \frac{p}{2} \log(\sigma^2) - \log|\phi I_p + X'X|^{\frac{1}{2}}. \quad (13)$$

علاوه بر این، اثر هم خطی ضرایب رگرسیونی با استفاده از توزیع پسین و پیشین از رابطه ذیل به دست می آید (لیندلی [۱۵]):

$$\vartheta_X(\beta|\phi) = H(\beta|\tau^2) - H_X(\beta|y, \sigma^2, \tau^2) \\ = \log|I_p + \phi^{-1}X'X|^{\frac{1}{2}}, \quad (14)$$

که رابطه (۱۴) با توجه به نسبت $\phi = \frac{\sigma^2}{\tau^2}$ حاصل می شود. در بخش بعدی، هم خطی را با کمک آنتروپی ماکسیم تعمیم یافته برطرف می کنیم.

۵ آنتروپی ماکسیم تعمیم یافته

آنتروپی ماکسیم تعمیم یافته^۴ بر اساس رویکرد آنتروپی ماکسیم کلاسیک جینز [۷] است و واضح است که در چارچوب آنتروپی ماکسیم کلاسیک، گشتاورها دقیق فرض شده اند. آنتروپی ماکسیم تعمیم یافته در اقتصاد و به ویژه در مدل های خطی به کار می رود که با مسئله هم خطی روبه رو هستند. این رویکرد به وسیله گولان و همکاران [۴] توسعه پیدا کرده است. آنتروپی ماکسیم تعمیم یافته بیشتر در مدل های پارامتری شده به کار می رود. بنا بر این می خواهیم β و ε (با فرض مستقل بودن خطاها) را از روش GME تحت مدل رگرسیونی

$$Y_{T \times 1} = X_{T \times K} \beta_{K \times 1} + \varepsilon_{T \times 1}$$

به دست آوریم. در این مدل، β ضریب رگرسیونی و ε عبارت خطا می باشند. فرض کنید z_k برداری $M \times 1$ برای پارامتر β_k

با توجه به $I_X(\beta|\sigma^2)$ و $I_{X^0}(\beta|\sigma^2)$ اندازه اطلاع $IL_X(\beta|\sigma^2)$ به صورت زیر تعریف می شود:

$$IL_X(\beta|\sigma^2) = I_{X^0}(\beta|\sigma^2) - I_X(\beta|\sigma^2) \\ = -\log|X'X|^{\frac{1}{2}} \\ = -\sum_{j=1}^p \log \lambda_j^{\frac{1}{2}}. \quad (9)$$

این اندازه اطلاع برای تشخیص هم خطی مفید است؛ چون بر حسب ویژه مقدارهای ماتریس $X'X$ است (بلسلی و همکاران [۲]). قابل ذکر است که ویژه مقدارهای کوچک نشان دهنده این است که بین متغیرهای پیشگو وابستگی خطی زیادی وجود دارد. یکی دیگر از راه های تشخیص هم خطی با استفاده از نظریه اطلاع، با کمک اطلاع کولبک-لیبلر است. برای این منظور فرض کنید

$$f_X(b|\beta, \sigma^2) = N(\beta, \sigma^2(X'X)^{-1})$$

$$f_{X^0}(b^0|\beta, \sigma^2) = N(\beta, \sigma^2 I_n)$$

باشد.

آن گاه اطلاع کولبک-لیبلر بین $f_X(b|\beta, \sigma^2)$ و $f_{X^0}(b^0|\beta, \sigma^2)$ عبارت است از

$$K(b : b^0) = K\{f_X(b|\beta, \sigma^2) : f_{X^0}(b^0|\beta, \sigma^2)\} \\ = \frac{1}{2} \{tr(X'X)^{-1} - \log|X'X|^{-1} - p\} \\ = \frac{1}{2} \left(\sum_{j=1}^p \frac{1}{\lambda_j} + \sum_{j=1}^p \log \lambda_j - p \right), \quad (10)$$

که $tr(X'X)^{-1}$ و $|X'X|$ برای تشخیص هم خطی مفید هستند (چون بر حسب ویژه مقدارهای ماتریس $X'X$ هستند). در نتیجه $K(b : b^0)$ اندازه زیان اطلاع ناشی از ماتریس غیر متعامد X است، همچنین

$$K(b : b^0) \geq 0$$

است و تساوی رخ خواهد داد اگر و فقط اگر ماتریس X متعامد باشد. اگر $K(b : b^0) \rightarrow \infty$ (وقتی $\lambda_j \rightarrow \infty$)، بین متغیرهای پیشگو هم خطی وجود دارد. اگر در فرضیات مسئله توزیع پیشین ضرایب رگرسیونی داده شده باشد می توانیم اثر هم خطی را به دست آوریم.

حال اگر فرض کنیم توزیع پیشین ضرایب رگرسیونی برابر با $\Pi(\beta|m, \tau^2) = N(m, \tau^2 I_p)$ باشد، آنتروپی پیشین ضرایب

^۴ maximum generalized entropy

3σ استفاده کرده‌اند. مقدار σ نیز همان انحراف معیار نمونه‌ای بردار (Y) است. در نتیجه مدل رگرسیونی به صورت

$$Y = X\beta + \varepsilon = XZP + \nu W,$$

است. بنا بر این احتمالات برآورد شده برای β و ε به صورت زیر خواهد بود:

$$\hat{p}_{km} = \frac{\exp(-z_{km} \sum_{t=1}^T x_{tk} \hat{\lambda}_t)}{\sum_{m=1}^M \exp(-z_{km} \sum_{t=1}^T x_{tk} \hat{\lambda}_t)},$$

$$m = 1, \dots, M, \quad k = 1, \dots, K,$$

$$\hat{w}_{tj} = \frac{\exp(-v_{tj} \hat{\lambda}_t)}{\sum_{j=1}^J \exp(-v_{tj} \hat{\lambda}_t)}, \quad j = 1, \dots, J, \quad t = 1, \dots, T.$$

در نهایت، مقادیر برآورد شده β و ε برابر خواهد بود با

$$\hat{\beta}_k \equiv \sum_{m=1}^M z_{km} \hat{p}_{km},$$

$$\hat{\varepsilon}_t \equiv \sum_{j=1}^J v_{tj} \hat{w}_{tj}.$$

همچنین λ_t ها ضرایب لاگرانژ می‌باشند که از رابطه زیر به دست می‌آیند:

$$\lambda_t = \text{Min} \left\{ \sum_{t=1}^T y_t \lambda_t + \sum_{k=1}^K \log \left(\sum_{m=1}^M \exp(-z_{km} \lambda_t x_{tk}) \right) \times \sum_{t=1}^T \lambda_t x_{tk} \right\} + \sum_{t=1}^T \log \left(\sum_{j=1}^J \exp(-\lambda_t v_{tj}) \right).$$

همان‌طور که ملاحظه می‌کنید، برای محاسبه $\hat{\beta}_k$ ، $\hat{w}_{tj}$ ، $\hat{p}_{km}$ و $\hat{\varepsilon}_t$ نیاز به محاسبه λ_t ها داریم. اما برای λ_t فرم بسته‌ای نداریم؛ بنا بر این برای محاسبه $\hat{\beta}_k$ ، $\hat{w}_{tj}$ ، $\hat{p}_{km}$ و $\hat{\varepsilon}_t$ از روش عددی استفاده می‌کنیم.

آنتروپی ماکسیمیم تعمیم‌یافته بر اساس آنتروپی ماکسیمیم کلاسیک به دست می‌آید؛ با این تفاوت که هیچ محدودیتی بر ضرایب و جمله خطای مدل رگرسیونی چندگانه وجود ندارد و هر دو پارامتر را با توجه به جمع‌پذیر بودن آنتروپی به روش عددی محاسبه می‌کنند. در روش آنتروپی ماکسیمیم، محدودیت‌ها به‌طور دقیق، تنها برای یک پارامتر ذکر می‌شوند. اگر در روش آنتروپی ماکسیمیم تعمیم‌یافته جداگانه برای هر یک از پارامترها محدودیت‌هایی مشخص شود، آنتروپی ماکسیمیم ضرایب و خطای مدل رگرسیونی به دست می‌آیند.

مثال ۱.۵. داده‌های جداول ۱ و ۲ مربوط به میزان محصول (بر حسب کیلوگرم) زمین‌های زراعی در سه سال متفاوت در فصل

تعریف شده باشد و همچنین اعضای بردار z_k از کوچک (z_{k1}) به بزرگ (z_{kM}) مرتب شده باشند. آن‌گاه برای هر پارامتر β_k بردار احتمال p_k ($\sum_{m=1}^M p_{km} = 1$)، وجود دارد به قسمی که

$$\beta_k = \sum_{m=1}^M p_{km} z_{km}, \quad k = 1, \dots, K, \quad M \geq 2.$$

به‌طور مشابه فرض می‌کنیم v_t برداری $1 \times J$ برای خطاها تعریف شده باشد و برای هر ε_t بردار احتمال w_t ($\sum_{j=1}^J w_{tj} = 1$)، وجود داشته باشد. بنا بر این تساوی زیر برقرار است:

$$\varepsilon_t = \sum_{j=1}^J w_{tj} v_j, \quad t = 1, \dots, T, \quad J \geq 2.$$

قابل ذکر است که β_k و ε_k متغیرهای تصادفی پیوسته‌اند.

برای منابع عدم قطعیت مستقل، آنتروپی جمع‌پذیر است (برای جزئیات بیشتر به کاپور و کواسان [۱۰] مراجعه شود). با فرض این که بردارهای احتمال p_k و w_t مستقل از یکدیگر باشند، می‌توان با حل مسئله بهینه‌سازی آنتروپی ماکسیمیم، پارامترها و خطاهای مجهول مدل رگرسیونی را برآورد کرد. بنا بر این روش آنتروپی ماکسیمیم تعمیم‌یافته به صورت زیر است:

$$\arg \text{Max}_{P,W} (H(P) + H(W)) =$$

$$\arg \text{Max}_{P,W} \left\{ - \sum_{k=1}^K \sum_{m=1}^M p_{km} \log p_{km} - \sum_{t=1}^T \sum_{j=1}^J w_{tj} \log w_{tj} \right\},$$

به قسمی که

$$y_t = \sum_{k=1}^K \sum_{m=1}^M z_{km} p_{km} x_{tk} + \sum_{j=1}^J v_j w_{tj}, \quad t = 1, \dots, T,$$

و

$$\sum_{m=1}^M p_{km} = 1, \quad k = 1, \dots, K,$$

$$\sum_{j=1}^J w_{tj} = 1, \quad t = 1, \dots, T.$$

برای تعیین کران‌های تکیه‌گاه خطاها، گولان و همکاران [۹] از قانون 3σ که توسط پوکلس‌هایم [۱۶] مطرح شده بود استفاده کردند. آن‌ها کران پایین را -3σ قرار داده و برای کران بالا از

در نتیجه مدل رگرسیونی چندگانه به صورت زیر است:

$$Y = 0.474568 + 0.906531X_1 + 81/33776X_2 + 0.147458X_3 + 47/98822.$$

در جدول ۴ (جدول برآورد پارامترها)، ضرایب و خطاهای مدل رگرسیونی برآورد شده و p -مقدار هر یک از ضرایب رگرسیونی، کمتر از ۰.۰۵ است، که نتیجه می شود، همه ضرایب، معنی دار هستند.

۶ نتیجه گیری

مسئله هم خطی یک مشکل جدی در مبحث رگرسیون است؛ زیرا با وجود هم خطی در مدل رگرسیونی، ضرایب مدل به درستی برآورد نمی شوند و دقت مدل کاهش می یابد. یکی از راه های برطرف کردن مشکل هم خطی، به کارگیری مدل رگرسیونی ستیغی می باشد. همچنین می توان با کمک نظریه اطلاع، از جمله آنتروپی ماکسیم تعمیم یافته، مشکل هم خطی را برطرف کرد. در این روش، خطاها و ضرایب رگرسیونی به روش های عددی به دست می آیند.

بهار است. هدف از جمع آوری این داده ها یافتن رابطه بین عواملی مانند میزان بارندگی (X_1)، دمای هوا (X_2)، دمای خاک (X_3) و رطوبت هوا (X_4) است، که کشاورزان در آن دخیل نیستند. در این مثال، مدل رگرسیونی عبارت است از

$$Y = 135/98056 + 23/20628X_1 + 72/18787X_2 + 15/23298X_3 - 21/18647X_4. \quad (15)$$

با استفاده از نرم افزار SAS مسئله هم خطی بین داده ها را بررسی کردیم و جدول همبستگی را به دست آوردیم (جدول ۳). در جدول ۳ می بینیم که بین X_2 و X_4 یک رابطه کاملاً خطی وجود دارد؛ یعنی همبستگی بین X_2 و X_4 برابر با یک است، که این نشان دهنده وجود هم خطی می باشد. بنا بر این با توجه به جدول ضرایب همبستگی می فهمیم که دمای خاک، تابع خطی از دمای هوا است. همچنین $K(b : b^\circ) = 6/89$ می باشد (با استفاده از نرم افزار R). می خواهیم با توجه به نظریه اطلاع، مدل رگرسیونی چندگانه را برای این داده ها به دست آوریم. اکنون با کمک آنتروپی ماکسیم تعمیم یافته و با استفاده از نرم افزار SAS، ضرایب مدل رگرسیونی را برآورد می کنیم.

جدول ۱. داده‌های میزان محصول زمین‌های زراعی

شماره مشاهده	X_1	X_2	X_3	X_4	Y
۱	۲۲/۶۸	۱۶/۱۷	۷۱/۹۱	۲۴/۲۵	۲۴۱۹/۳۲
۲	۱۴/۴۶	۱۴/۳۴	۷۶/۰۸	۲۱/۵۱	۲۲۰۴/۵۴
۳	۱۲/۷۷	۱۵/۱۸	۶۹/۰۵	۲۲/۷۷	۲۰۹۲/۶۲
۴	۲۷/۷۳	۱۵/۳۷	۷۶/۸۶	۲۳/۰۵	۲۵۶۶/۰۶
۵	۱۲/۹۷	۱۵/۵۰	۷۴/۲۵	۲۳/۲۵	۲۲۰۱/۴۸
۶	۲۷/۷۴	۱۵/۴۸	۷۷/۳۱	۲۳/۲۲	۲۶۱۷/۵۸
۷	۱۶/۱۴	۱۴/۱۹	۷۳/۲۱	۲۱/۲۹	۲۱۷۸/۳۲
۸	۱۹/۵۲	۱۶/۵۷	۸۰/۳۳	۲۴/۸۵	۲۴۸۳/۶۹
۹	۲۵/۹۵	۱۴/۹۱	۷۵/۷۴	۲۲/۳۶	۲۴۷۸/۸۹
۱۰	۱۹/۳۳	۱۴/۹۲	۷۲/۲۲	۲۲/۳۷	۲۲۸۱/۹۲
۱۱	۲۱/۴۳	۱۴/۶۰	۷۳/۱۹	۲۱/۸۹	۲۳۲۳/۰۲
۱۲	۳۰/۵۴	۱۵/۹۵	۷۲/۶۸	۲۳/۹۳	۲۵۹۴/۰۲
۱۳	۱۵/۸۶	۱۵/۲۳	۷۵/۵۶	۲۲/۸۳	۲۲۸۵/۱۸
۱۴	۳۳/۲۷	۱۵/۴۷	۷۴/۲۰	۲۳/۲۱	۲۶۷۴/۷۳
۱۵	۳۳/۸۷	۱۶/۰۱	۷۶/۱۹	۲۴/۰۱	۲۷۳۰/۳۳
۱۶	۲۵/۷۳	۱۴/۸۱	۷۴/۸۶	۲۲/۲۱	۲۴۹۱/۷۷
۱۷	۱۹/۱۳	۱۴/۷۲	۷۳/۰۹	۲۲/۰۷	۲۲۸۹/۲۳
۱۸	۲۷/۸۵	۱۵/۶۶	۷۷/۱۸	۲۳/۴۸	۲۵۸۷/۹۵
۱۹	۱۲/۷۳	۱۵/۲۲	۷۳/۱۲	۲۲/۸۳	۲۱۶۰/۶۴
۲۰	۲۴/۴۶	۱۵/۳۶	۷۴/۵۶	۲۳/۰۳	۲۴۶۰/۷۴
۲۱	۲۰/۹۷	۱۶/۱۲	۷۵/۴۷	۲۴/۱۸	۲۴۱۹/۹۶
۲۲	۲۰/۴۲	۱۵/۰۵	۷۱/۲۲	۲۲/۵۸	۲۳۰۰/۲۰
۲۳	۱۶/۶۹	۱۵/۲۸	۷۲/۶۰	۲۲/۹۲	۲۲۴۶/۱۷
۲۴	۲۳/۸۰	۱۵/۶۹	۷۱/۷۵	۲۳/۵۴	۲۴۲۲/۹۰
۲۵	۲۴/۲۶	۱۵/۳۵	۷۵/۵۱	۲۳/۰۱	۲۴۷۸/۶۴
۲۶	۲۷/۵۷	۱۵/۶۴	۷۴/۲۸	۲۳/۴۷	۲۵۲۵/۹۷
۲۷	۲۴/۶۸	۱۵/۳۵	۷۰/۹۷	۲۳/۰۲	۲۴۱۴/۲۶
۲۸	۳۱/۷۷	۱۷/۳۵	۷۱/۵۶	۲۶/۰۲	۲۶۵۲/۸۰
۲۹	۲۰/۴۳	۱۴/۹۲	۷۰/۱۴	۲۲/۳۷	۲۲۷۹/۹۹
۳۰	۲۳/۷۱	۱۴/۸۹	۷۵/۵۳	۲۲/۳۵	۲۴۳۴/۷۱
۳۱	۱۷/۹۳	۱۵/۱۳	۷۳/۹۳	۲۲/۷۱	۲۲۹۳/۴۰

جدول ۲ (ادامه). داده‌های میزان محصول زمین‌های زراعی

شماره مشاهده	X_1	X_2	X_3	X_4	Y
۳۲	۲۷/۲۲	۱۵/۳۷	۷۸/۴۵	۲۳/۰۵	۲۵۶۵/۸۳
۳۳	۲۲/۴۶	۱۶/۳۰	۷۳/۳۵	۲۴/۴۶	۲۴۳۹/۰۹
۳۴	۱۸/۶۸	۱۵/۱۶	۷۴/۴۵	۲۲/۷۴	۲۳۱۶/۶۹
۳۵	۱۹/۷۶	۱۶/۲۷	۷۷/۲۹	۲۴/۴۲	۲۴۲۲/۹۰
۳۶	۱۵/۲۱	۱۶/۰۸	۷۵/۴۷	۲۴/۱۱	۲۲۹۰/۱۸
۳۷	۱۸/۳۶	۱۵/۱۵	۷۳/۷۳	۲۲/۷۳	۲۳۱۸/۰۸
۳۸	۲۸/۶۹	۱۶/۲۲	۷۳/۰۱	۲۴/۳۳	۲۵۷۷/۶۲
۳۹	۲۳/۶۲	۱۶/۱۰	۷۲/۴۴	۲۴/۱۴	۲۴۳۹/۰۰
۴۰	۲۲/۸۱	۱۵/۸۲	۷۴/۱۴	۲۳/۷۳	۲۴۲۶/۸۰
۴۱	۲۸/۵۰	۱۶/۵۴	۶۹/۲۵	۲۴/۸۰	۲۵۱۶/۸۷
۴۲	۱۶/۳۳	۱۶/۳۶	۷۹/۵۹	۲۴/۵۳	۲۳۷۷/۸۹
۴۳	۱۶/۲۷	۱۵/۸۲	۷۶/۱۵	۲۳/۷۲	۲۳۲۲/۹۹
۴۴	۲۴/۱۲	۱۵/۷۳	۷۲/۰۸	۲۳/۶۰	۲۴۲۶/۸۱
۴۵	۲۹/۰۶	۱۷/۱۴	۷۷/۷۸	۲۵/۷۰	۲۶۷۹/۲۶

جدول ۳. همبستگی مثال ۱

متغیر	X_1	X_2	X_3	X_4	Y
X_1	۱/۰۰۰۰	۰/۳۸۵۹	۰/۰۵۰۴	۰/۳۸۶۱	۰/۹۴۲۸
X_2	۰/۳۸۵۹	۱/۰۰۰۰	۰/۱۵۸۸	۱/۰۰۰۰	۰/۵۵۴۰
X_3	۰/۰۵۰۴	۰/۱۵۸۸	۱/۰۰۰۰	۰/۱۵۸۲	۰/۳۲۸۰
X_4	۰/۳۸۶۱	۱/۰۰۰۰	۰/۱۵۸۲	۱/۰۰۰۰	۰/۵۵۴۰
Y	۰/۹۴۲۸	۰/۵۵۴۰	۰/۳۲۸۰	۰/۵۵۴۰	۱/۰۰۰۰

جدول ۴. برآورد پارامترهای مثال ۱

متغیر	برآورد پارامتر	انحراف معیار	مقدار t	$Pr > t $
X_1	۰/۹۰۶۵۳۱	۰/۱۳۵۳	۶/۷۰	$< ۰/۰۰۰۱$
X_2	۸۱/۳۳۷۷۶	۰/۶۱۶۸	۱۳۱/۸۸	$< ۰/۰۰۰۱$
X_3	۰/۱۴۷۴۵۸	۰/۰۱۷۹	۸/۲۴	$< ۰/۰۰۰۱$
X_4	۴۷/۹۸۸۲۲	۰/۳۶۴۰	۱۳۱/۸۳	$< ۰/۰۰۰۱$
عرض از مبدأ	۰/۴۷۴۵۶۸	۰/۰۴۵۳	۱۰/۴۷	$< ۰/۰۰۰۱$

مراجع

- [۱] وحیدیان و بزرگ‌نیا (۱۳۷۲). نظریه کنترل و کنترل بهینه. دانشگاه فردوسی مشهد.
- [2] Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Observation and Sources of Collinearity*. Wiley, New York.
- [3] Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley, New York.
- [4] Golan, A., Judge, G., and Miller, J. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Wiley, New York.
- [5] Hartley, R. V. L. (1928). Transmission of information. *Bell System Technical Journal*, **7**, (3), 535-563.
- [6] Horel, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, **58**, 54-59.
- [7] Jaynes, E. T. (1968). On the rational of maximum entropy methods. *Proceedings of Institute of Electrical and Electronics Engineers*, **70**, 939-952.
- [8] Jeffreys, H. (1948). An invariant form for the prior probability in estimating problems. *Proceedings of the Royal Society of London*, **186**, 453-461.
- [9] Kagan, A. M., Linnik, Y. V., and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.
- [10] Kupur, J. N. and Kesavan, H. K. (1992). *Entropy Optimization Principles with Applications*. Academic Press, London.
- [11] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.
- [12] Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to Linear Regression Analysis*. Wiley, New York.
- [13] Nyquist, H. (1924). Certain factors affecting telegraph speed. *Bell System Technical Journal*, **3**, 126-147.
- [14] Leamer, E. E. (1973). Multicollinearity: A Bayesian interpretation. *Review of Economics and Statistics*, **55**, 371-380.
- [15] Lindley, D. V. (1956). On a measure of information provided by an experiment. *Annals of Mathematical Statistics*, **27**, 986-1005.
- [16] Pukelsheim, F. (1994). The three-sigma rule. *The American Statistician*, **4**, 88-91.
- [17] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423.