سیزدهمین کنفرانس آمار ایران
۲-۴ شهریور ۱۳۹۵

آزمون معنی داری ضرایب و انتخاب مدل بهینه برای رگرسیون با کمک نظریه اطلاعات

فائزه شکیبا[۱]، غلامرضا محتشمی برزادران [۲]

[۱]دانشگاه فردوسی مشهد

[۲]دانشگاه فردوسی مشهد

چکیده: در این مقاله ابتدا مفاهیم مقدماتی نظریه اطلاعات به طور مختصر بیان می شود. در ادامه اصل آنتروپی ماکسیمم مطرح می شود و همچنین نحوه به دست آوردن آنتروپی ماکسیمم متغیر پاسخ، برآورد ضرایب و واریانس، آزمون معنی داری ضرایب و انتخاب مدل بهینه در مدل رگرسیون چندگانه با کمک نظریه اطلاعات عنوان می شود.

واژه‌های کلیدی: آنتروپی، آنتروپی ماکسیمم، اطلاع کولبک– لیبلر، آنتروپی نسبی مینیمم و آنتروپی نسبی مینیمم بیزی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 99X99، 99X99، 99X99.

---

۱ مقدمه

شروع تحقیق در مورد نظریه اطلاعات اولین بار در سال ۱۹۲۴ توسط هری نیکوئیست [۱] در مقاله ای به نام " عوامل خاصی که سرعت تلگراف را تحت تاثیر قرار می دهند " انجام شد. کار مهم دیگر در این زمان مقاله " انتقال اطلاعات " در سال ۱۹۲۸ توسط هارتلی [۲] بود که اولین پایه های ریاضی نظریه اطلاعات را بنا گذاشت. تولد واقعی نظریه اطلاعات را می توان به مقاله " نظریه ریاضی مخابرات " توسط کلود شانون " [۳] در سال ۱۹۴۸ نسبت داد. کنکاش و توسیع بحث آنتروپی از سال ۱۹۴۸ تا کنون تحولات بسیاری را موجب شده است که از آن جمله بحث آنتروپی ماکسیمم یکی از آنهاست. آنتروپی ماکسیمم که توسط جینز [۴] در سال ۱۹۶۸ مطرح شد، اشاره به این مطلب دارد که در انتخاب توزیع برای متغیر فقط و فقط باید از اطلاعاتی که داریم استفاده کنیم. به این معنی که باید توزیعی را در نظر بگیریم که تحت اطلاعات داده شده بیشترین عدم قطعیت یا کمترین تصادفی بودن را داشته باشد، چرا که اگر از توزیعی با عدم قطعیت

---

[۲]فائزه شکیبا : shakiba.faeze@yahoo.com

A-10-189-1

Nyquist[۱]

Hartley[۲]

Shannon [۳]

Jaynes[۴]

کمتر استفاده کنیم به این معنی است که از اطلاعاتی که به ما داده نشده استفاده کرده ایم. در طی چهار دهه اخیر رابطه بین آنتروپی و رگرسیون گسترش پیدا کرده است. کولبک و روزن بلات [۵] (۱۹۵۷) در زمینه گسترش رگرسیون و نظریه آنتروپی پیشگام بوده اند. آنتروپی برای پیش بینی، برآورد و آزمون معنی داری پارامترهای رگرسیون نقش دارد که در این مقاله این موضوع مورد بررسی قرار می گیرد.

۲ مفاهیم مقدماتی مرتبط با نظریه اطلاعات

به طور کلی نظریه اطلاعات مدلی ریاضی از شرایط و عوامل موثر در انتقال و پردازش داده ها و اطلاعات فراهم می آورد. آنتروپی اطلاعات که به آنتروپی شانون هم شناخته می شود در واقع میزان تصادفی بودن یک رخداد را به صورت یک نسخه ریاضی گزارش می کند. اگر X متغیر تصادفی پیوسته با تابع چگالی $f(x)$ باشد آنگاه آنتروپی شانون عبارت است از:

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

همچنین کولبک-لیبلر [۶] در سال ۱۹۵۱ اندازه اطلاع کولبک- لیبلر را معرفی کردند که این اندازه را آنتروپی نسبی یا اطلاع واگرا هم گویند و چون که این اندازه اختلاف بین دو توزیع از متغیر تصادفی را به دست می آورد، دارای اهمیت می باشد. اگر X و Y دو متغیر تصادفی پیوسته با تابع چگالی های $f(x)$ و $g(x)$ باشند، اطلاع کولبک- لیبلر برابر با $K(f:g) = \int_{-\infty}^{\infty} f(x)\log\frac{f(x)}{g(x)}dx$ است.

مثال ۱.۲. اگر $f \sim N(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f)$ و $g \sim N(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ باشند، آنگاه،

$$K(f:g) = \frac{۱}{۲}(tr(\boldsymbol{\Sigma}_f\boldsymbol{\Sigma}_g^{-۱}) - n) - \frac{۱}{۲}\log\frac{|\boldsymbol{\Sigma}_f|}{|\boldsymbol{\Sigma}_g|} + \frac{۱}{۲}(\boldsymbol{\mu}_f - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_g^{-۱}(\boldsymbol{\mu}_f - \boldsymbol{\mu}_g).$$

توجه داشته باشیم tr اثر ماتریس $\boldsymbol{\Sigma}_f\boldsymbol{\Sigma}_g^{-۱}$ می باشد.

۳ آنتروپی ماکسیمم

آنتروپی ماکسیمم یکی از مباحث مهم در نظریه اطلاعات است که در این بخش می خواهیم آنتروپی ماکسیمم شانون را به دست آوریم. برای این منظور ابتدا قضیه ۱.۳ را مطرح می کنیم.

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 ماکسیمم می کند از تساوی زیر به دست می آید (وحیدیان و بزرگ نیا [۷] ۱۳۷۲):

$$\frac{\partial F}{\partial f(x)} - \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial f'(x)}\right) = ۰.$$

به کمک قضیه ۱.۳، کاگان [۸] و همکاران قضیه زیر را معرفی کردند که با کمک این قضیه می توان آنتروپی ماکسیمم شانون را برای متغیر تصادفی X به دست آورد.

[۵] Kullback and Rosenblatt

[۶] Kullback and Leibler

[۷] Vahidian and Bozorgnia

[۸] Kagan

قضیه ۲.۳. هرگاه X متغیر تصادفی باشد که به ازای هر x متعلق به تکیه گاه D، $f(x) > 0$ و همچنین توابع $r_i(x), i = 1, ..., k$ طوری باشند که $\int_D f(x) = 1$ ، $\int_D f(x)r_i(x) = \alpha_i, i = 1, ..., k$ و α_i ها ثابت باشند، آنگاه توزیع دارای آنتروپی ماکسیمم به فرم نمایی $f^*(x) = e^{\lambda_0+\lambda_1 r_1(x)+...+\lambda_k r_k(x)}$ می باشد که $\lambda_0, \lambda_1, ..., \lambda_k$ به کمک شرایط ذکر شده به دست می آیند (کاگان و همکاران ۱۹۷۳).

چون در آنتروپی ماکسیمم تابع f' را نداریم بنابراین تابع زیر علامت انتگرال تابعی از x و $f(x)$ است و فقط $\frac{\partial F}{\partial f(x)}$ را باید حل کنیم.

به عنوان مثال اگر برای متغیر تصادفی نامنفی $EX = a$ آنتروپی ماکسیمم به شکل نمایی و اگر $EX = a$ و $EX^2 = b$ آنتروپی ماکسیمم به صورت توزیع نرمال و در نهایت هیچ شرطی ذکر نشود، آنتروپی ماکسیمم دارای توزیع یکنواخت خواهد بود.

۴ آنتروپی ماکسیمم برای متغیر پاسخ مدل رگرسیون خطی

مدل رگرسیون خطی چندگانه به فرم $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$ می باشد که $\boldsymbol{Y}$ بردار مشاهدات $n \times 1$، $\mathbf{X}$ نشان دهنده ماتریس رگرسیون از مرتبه $n \times p$، $\boldsymbol{\beta}$ بردار $p \times 1$ از ضرایب رگرسیون و ε بردار تصادفی $n \times 1$ از عبارت های خطا می باشد ($\varepsilon \sim N(0, \sigma^2\boldsymbol{I}_n)$).

حال با توجه به قضیه ۱۰۳ ، آنتروپی ماکسیمم متغیر پاسخ مدل رگرسیون چندگانه تحت محدودیت های $E(\boldsymbol{Y}) = \mathbf{X}\boldsymbol{\beta}$ و $Var(\boldsymbol{Y}) = \sigma^2\mathbf{I}_n$ برابر با $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$ می باشد.

۵ برآورد و پیش بینی رگرسیون

مجموعه ای از مشاهدات $y_1, ..., y_n$ که نمایشی از توزیع n متغیره $f(\boldsymbol{y})$ می باشند را در اختیار داریم، هدف برآورد $\boldsymbol{\theta}$ از آنتروپی ماکسیمم متغیر پاسخ مدل رگرسیون می باشد ($\boldsymbol{\theta}$ یک بردار p تایی از ضرایب است). یکی از راه های برآورد پارامتر $\boldsymbol{\theta}$ بکارگیری آنتروپی نسبی مینیمم (MDI) است. برآورد آنتروپی نسبی مینیمم پارامتر $\boldsymbol{\theta}$ برحسب اطلاع کولبک- لیبلر بین $f(\boldsymbol{y})$ و $f^*(\boldsymbol{y}, \tilde{\boldsymbol{\theta}})$ عبارتست از (صوفی [۹] ۱۹۹۶)،

$$\tilde{\boldsymbol{\theta}}_{MDI} = arg_{\boldsymbol{\theta}} Min K\{f(\boldsymbol{y}) : f^*(\boldsymbol{y}, \tilde{\boldsymbol{\theta}})\},$$

همچنین می توان پارامتر $\boldsymbol{\theta}$ را از روش آنتروپی نسبی مینیمم بیزی $(MDIB)$ به دست آورد که آنتروپی نسبی مینیمم بیزی پارامتر $\boldsymbol{\theta}$ بر حسب متوسط اطلاع کولبک- لیبلر از رابطه (صوفی ۱۹۹۶)

$$\tilde{\boldsymbol{\theta}}_{MDIB} = arg_{\boldsymbol{\theta}} Min E\{\overline{K}\{f(\boldsymbol{y}) : f^*(\boldsymbol{y}, \tilde{\theta})\}|y\}, \tag{۱.۵}$$

حاصل می شود. قابل ذکر است که متوسط اطلاع کولبک-لیبلر بین $f(\boldsymbol{y})$ و $f^*(\boldsymbol{y}, \boldsymbol{\theta})$ بر حسب $\tilde{\boldsymbol{\theta}}$، با توجه به اینکه $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ می باشد، به صورت $\overline{K}\{f(\boldsymbol{y}) : f^*(\boldsymbol{y}, \tilde{\boldsymbol{\theta}})\} \equiv \frac{1}{n}K\{f(\boldsymbol{y}) : f^*(\boldsymbol{y}, \tilde{\theta})\}$, تعریف می شود.

مثال ۱.۵. قبلا اشاره کردیم که آنتروپی ماکسیمم متغیر پاسخ مدل رگرسیون تحت محدودیت های $E(\boldsymbol{Y}) = \mathbf{X}\boldsymbol{\beta}$ و $Var(\boldsymbol{Y}) = \sigma^2\boldsymbol{I}_n$ برابر با $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$ می باشد. در این مثال می خواهیم برآورد آنتروپی نسبی مینیمم بیزی $\boldsymbol{\beta}$ و σ^2 را به

---
[۹] Soofi

دست آوریم. برای این منظور تابع دلخواه $f(y,\boldsymbol{\mu},\boldsymbol{\Omega})$ را در نظر می گیریم و فرض می کنیم $\boldsymbol{\Omega}=\omega^2\boldsymbol{I}_n$ باشد. با توجه به رابطه (۱۰۵) ، برابری های زیر برقرارند:

$$\tilde{\beta}_{MDIB}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\mu}|\boldsymbol{y}), \tag{۲۰۵}$$

$$\tilde{\sigma}^2_{MDIB}=E(\omega^2|\boldsymbol{y})+\frac{1}{n}E(\boldsymbol{\mu}'\boldsymbol{\mu}|\boldsymbol{y})+\frac{1}{n}E(\boldsymbol{\mu}'|\boldsymbol{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\mu}|\boldsymbol{y}). \tag{۳۰۵}$$

در رابطه (۲۰۵)، $E(\boldsymbol{\mu}|\boldsymbol{y})$، میانگین شرطی بردار $\boldsymbol{\mu}$ و در رابطه (۳۰۵)، $E(\boldsymbol{\mu}'\boldsymbol{\mu})$ میانگین شرطی بردار $\boldsymbol{\mu}'\boldsymbol{\mu}$ می باشند.

وقتی که هیچ اطلاعی درباره توزیع پیشین نداشته باشیم برآوردهای آنتروپی نسبی مینیمم بیزی روابط (۲۰۵) و (۳۰۵) تقریبا معادل با برآوردگرهای درستنمایی ماکسیمم β و σ^2 تحت مدل آنتروپی ماکسیمم متغیر پاسخ می باشند.

۶ آزمون های آنتروپی نسبی مینیمم بیزی

در حالت کلی فرض کنید رابطه $\boldsymbol{Y}=\mathbf{X}\beta+\varepsilon$ برقرار است که $\boldsymbol{Y}$ و ε بردارهایی $n\times 1$، $(\varepsilon\sim N(0,\sigma^2\boldsymbol{I}_n))$، $\mathbf{X}$ ماتریس تمام رتبه $n\times p$ و β برداری $p\times 1$ می باشند.

برای آزمون

$$\begin{cases} H_0: & \beta=\beta^0 \\ H_1: & \beta\neq\beta^0 \end{cases}$$

با کمک آنتروپی، نیاز به آنتروپی ماکسیمم متغیر پاسخ که به شکل $f^*(\boldsymbol{y},\beta,\sigma^2)=N(\mathbf{X}\beta,\sigma^2\boldsymbol{I}_n)$، و اطلاع کولبک- لیبلر که در مثال ۱۰۲ به دست آورده ایم، داریم. بنابراین اطلاع کولبک-لیبلر بین $f^*_{\boldsymbol{\beta}}$ و $f^*_{\boldsymbol{\beta}^0}$ عبارتست از،

$$K(f^*_{\boldsymbol{\beta}}:f^*_{\boldsymbol{\beta}^0},\sigma^2)=\frac{(\beta-\beta^0)'\mathbf{X}'\,\mathbf{X}(\beta-\beta^0)}{2\sigma^2}.$$

در نتیجه آماره آزمون برای آزمون فرضیه صفر $H_0:\beta=\beta^0$ در مقابل فرضیه $H_1:\beta\neq\beta^0$ برابر با اطلاع جفریز می باشد که از رابطه $J(f^*_{\boldsymbol{\beta}}:f^*_{\boldsymbol{\beta}^0},\sigma^2)=K(f^*_{\boldsymbol{\beta}}:f^*_{\boldsymbol{\beta}^0},\sigma^2)+K(f^*_{\boldsymbol{\beta}^0}:f^*_{\boldsymbol{\beta}},\sigma^2)$ به دست می آید. بنابراین آماره آزمون برابر با $J(f^*_{\boldsymbol{\beta}}:f^*_{\boldsymbol{\beta}^0},\sigma^2)=\dfrac{(\beta-\beta^0)'\mathbf{X}'\mathbf{X}(\beta-\beta^0)}{\sigma^2}$ است. در حالت خاص اگر $\beta^0=0$ باشد، آنگاه $\widehat{J}(f^*_{\boldsymbol{\beta}}:f^*_{\boldsymbol{\beta}^0},\sigma^2)=\dfrac{(\beta)'\mathbf{X}'\mathbf{X}(\beta)}{\widehat{\sigma}^2}$ که $\widehat{\sigma}^2$ همان برآورد آنتروپی نسبی مینیمم بیزی σ^2 است (صوفی ۱۹۹۶). قبلا اشاره کردیم که برآورد آنتروپی نسبی مینیمم بیزی β و σ^2 در معادله رگرسیون وقتی هیچ اطلاعی درباره توزیع پیشین نداشته باشیم معادل با برآوردگرهای درستنمایی ماکسیمم β و σ^2 تحت مدل آنتروپی ماکسیمم متغیر پاسخ می باشند. یعنی روابط $\tilde{\beta}_{MDIB}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{Y}=\boldsymbol{b}$ و $\tilde{\sigma}^2_{MDIB}\approx\frac{1}{n}(\boldsymbol{Y}-\mathbf{X}\boldsymbol{b})'(\boldsymbol{Y}-\mathbf{X}\boldsymbol{b})=\widehat{\sigma}^2$ برقرارند.

قابل ذکر است که $n\widehat{\sigma}^2=(n-p)s^2$ در نتیجه $s^2=\dfrac{(\boldsymbol{Y}-\mathbf{X}\boldsymbol{b})'(\boldsymbol{Y}-\mathbf{X}\boldsymbol{b})}{n-p}$. همچنین توجه داریم که روابط $\dfrac{(\beta)'\mathbf{X}'\mathbf{X}(\beta)}{p}\sim\chi^2_p$، و $\dfrac{(\boldsymbol{Y}-\mathbf{X}\boldsymbol{b})'(\boldsymbol{Y}-\mathbf{X}\boldsymbol{b})}{n-p}\sim\chi^2_{n-p}$ وجود دارند و آماره آزمون ($\widehat{J}(f^*_{\boldsymbol{\beta}}:f^*_{\boldsymbol{\beta}^0},\sigma^2)\sim pF_{p,n-p}$) از توزیع F چندگانه با درجه آزادی p و $n-p$ پیروی می کند. اگر $|\widehat{J}(f^*_{\boldsymbol{\beta}},f^*_{\boldsymbol{\beta}^0},\sigma^2)|<pF_{p,n-p}$ فرضیه H_0 را نمی توان رد کرد.

مثال ۱.۶. داده های جدول ۱ مربوط به میزان مهاجرت در کشور ایتالیا بین سال های ۱۹۷۸ تا ۲۰۰۴ می باشد (اورین و تونا [10]

[10] Evren and Tuna

(۲۰۱۲).

جدول ۱: داده های میزان مهاجرت کشور ایتالیا بین سال های ۲۰۰۴-۱۹۷۸

سال	خالص مهاجرت Y	سرانه (X_1)	نسبت بیکاری (X_2)	کل جمعیت (X_3)
۱۹۷۸	۰/۱	۵/۴۰۶	۷/۲	۵۵۴۴۶
۱۹۷۹	۰/۱	۶/۷۳۵	۷/۷	۵۵۶۰۲
۱۹۸۰	۰/۱	۸/۱۴۸	۷/۶	۵۵۶۵۷
۱۹۸۱	−۰/۳	۷/۳۴۴	۷/۹	۵۵۷۷۴
۱۹۸۲	۱/۹	۷/۲۸۱	۸/۵	۵۵۹۹۵
۱۹۸۳	۲/۴	۷/۵۴۷	۹/۴	۵۶۲۲۸
۱۹۸۴	۱/۶	۷/۴۵۷	۱۰/۱	۵۶۳۴۴
۱۹۸۵	−۰/۵	۷/۶۹۹	۱۰/۳	۵۶۴۹۸
۱۹۸۶	−۰/۴	۱۰/۹۰۲	۱۱/۱	۵۶۵۷۶
۱۹۸۷	−۰/۲	۱۳/۷۱۵	۱۲	۵۶۶۶۴
۱۹۸۸	۰	۱۵/۷۱	۱۲	۵۶۷۶۳
۱۹۸۹	۰/۲	۱۵/۷۸۸	۱۲	۵۶۸۳۷
۱۹۹۰	۰/۲	۱۹/۹۸۳	۱۱/۴	۵۶۷۳۷
۱۹۹۱	۰/۱	۲۱/۰۵۹	۱۰/۹	۵۶۷۶۰
۱۹۹۲	۳/۲	۲۲/۲۸۶	۱۰/۵	۵۶۸۵۹
۱۹۹۳	۳/۲	۱۷/۹۶۵	۹/۷	۵۶۴۴۲
۱۹۹۴	۲/۶	۱۸/۵۴	۱۰/۶	۵۶۶۲۳
۱۹۹۵	۱/۶	۱۹/۸۰۹	۱۱/۲	۵۶۷۴۵
۱۹۹۶	۲/۶	۲۲/۱۵۲	۱۱/۲	۵۶۸۲۶
۱۹۹۷	۲/۲	۲۰/۹۵۷	۱۱/۳	۵۶۹۴۱
۱۹۹۸	۱/۶	۲۱/۳۸۶	۱۱/۳	۵۷۰۴۰
۱۹۹۹	۱/۸	۲۱/۰۹۶	۱۱	۵۷۰۷۸
۲۰۰۰	۳/۱	۱۹/۲۶۹	۱۰/۲	۵۷۱۸۹
۲۰۰۱	۲/۲	۱۹/۶۰۹	۹/۱	۵۷۳۴۸
۲۰۰۲	۶/۱	۲۱/۳۲۶	۸/۶	۵۷۴۷۴
۲۰۰۳	۱۰/۶	۱۶/۱۶۴	۸/۵	۵۷۴۷۸
۲۰۰۴	۷/۹	۲۹/۶۹۹	۸/۱	۵۷۵۵۳

در این مثال می خواهیم بدانیم که کدامیک از عوامل ذکر شده بر میزان مهاجرت اثر دارد. بنابراین آزمون فرضیه صفر $H_\circ : \beta' = \begin{bmatrix} -۲۱۰/۸۴ & ۰/۰۰۶ & ۰ & ۳/۹۵ \end{bmatrix}$ را در مقابل $H_۱ : \beta^{\circ\prime} = \begin{bmatrix} -۲۱۰/۸۴ & ۰/۰۰۶ & -۱/۱۳ & ۳/۹۵ \end{bmatrix}$ در نظر می گیریم. در نتیجه $K\{f^*(\boldsymbol{y},\boldsymbol{\beta},\sigma^۲) : f^*(\boldsymbol{y},\boldsymbol{\beta}^\circ,\sigma^۲)\} = ۸۷۶$ و $J\{f^*(\boldsymbol{y},\boldsymbol{\beta},\sigma^۲) : f^*(\boldsymbol{y},\boldsymbol{\beta}^\circ,\sigma^۲)\} = ۱۷۵۲$

می باشند. با توجه به مقدار بزرگ آماره آزمون، نتیجه می گیریم فرضیه صفر رد می شود و یک رابطه معنی داری در مدل رگرسیون وجود دارد. یعنی اینکه نسبت بیکاری بر مهاجرت اثر دارد و عامل مهمی بر میزان مهاجرت است.

همچنین در آزمون های کلاسیک اشاره کردیم که $T \sim t_{n-2}$ آنگاه $T^2 \sim F_{1,n-2}$ است. حال می خواهیم در آزمون های بیزی به رابطه ای که بین توزیع F و توزیع t چند متغیره (T^2 هتلینگ [11]) وجود دارد، دست پیدا کنیم.

فرض کنیم ماتریس $\boldsymbol{M}_{p\times p}$ را بتوان به صورت $\boldsymbol{M} = \mathbf{X}'\mathbf{X}$ نوشت که $\mathbf{X}$ ماتریسی از مرتبه $n \times p$ و $\mathbf{X} \sim N(0, \boldsymbol{\Sigma})$ باشد، آنگاه می گویند $\boldsymbol{M}$ دارای توزیع ویشارت با ماتریس مقیاس $\boldsymbol{\Sigma}$ و درجه آزادی n است، می نویسیم $\boldsymbol{M} \sim W_p(\boldsymbol{\Sigma}, n)$. وقتی $\boldsymbol{\Sigma} = \boldsymbol{I}_p$، می گویند که توزیع به صورت استاندارد است.

قابل ذکر است که اگر $p = 1$ و $\boldsymbol{\Sigma} = \sigma^2$، آنگاه $\boldsymbol{M} \sim W_1(\sigma^2, n)$. یعنی توزیع $W_1(\sigma^2, n)$ همانند توزیع $\sigma^2\chi^2_n$ می باشد.

تعریف ۲.۶. اگر α را بتوانیم به صورت $m\boldsymbol{d}'\boldsymbol{M}^{-1}\boldsymbol{d}$ بنویسیم که در آن $\boldsymbol{d} \sim N(0, \boldsymbol{I})$ و $\boldsymbol{M} \sim W_p(\boldsymbol{I}, n)$، $\boldsymbol{M}$ و $\boldsymbol{d}$ از هم مستقل هستند، آنگاه می گوییم که α دارای توزیع T^2 هتلینگ با پارامترهای p و n می باشد ($\alpha \sim T^2(p, n)$).

قضیه ۳.۶. اگر $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ و $\boldsymbol{M} \sim W_p(\boldsymbol{\Sigma}, n)$ آنگاه (ماردیا [12] و همکاران ۱۹۷۹)

$$n(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{M}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim T^2(p, n).$$

مربع توزیع t یک متغیره با درجه آزادی n، دارای توزیع $T^2(1, n)$ است. به عبارت دیگر توزیع $F_{1,n}$ و توزیع $T^2(1, n)$ یکسان اند. در نتیجه آماره آزمون بیزی دارای توزیع T^2 هتلینگ می باشد.

۷ مدل های بهینه

با کمک نظریه اطلاعات می خواهیم مدل بهینه را برای مدل رگرسیون پیدا کنیم. مدل M_{k^*} را بهینه گویند، هرگاه رابطه $\tilde{K}(f(\boldsymbol{y}) : f^*_{k^*}(\boldsymbol{y}, \tilde{\boldsymbol{\theta}}_{k^*})) \leq \tilde{K}(f(\boldsymbol{y}) : f^*_k(\boldsymbol{y}, \tilde{\boldsymbol{\theta}}_k))$ برقرار باشد. در آن $\tilde{\boldsymbol{\theta}}$ برآورد پارامتر مدل و $\tilde{K}$ برآورد اطلاع کولبک–لیبلر می باشند. معیار دیگری برای بهینه بودن مدل معیار اطلاع F است که به شکل $(M_k \leq M_l)$, $FIC(k) = \dfrac{2\widehat{K}(f^*_{\boldsymbol{\beta}_k} : f^*_{\boldsymbol{\beta}_l}, \sigma^2)}{p_l - p_k}$ می باشد. اگر رابطه $2\widehat{K}(f^*_{\boldsymbol{\beta}_k} : f^*_{\boldsymbol{\beta}_l}, \sigma^2) = 2K(f^*_{\boldsymbol{b}_k} : f^*_{\boldsymbol{b}_l}, \widehat{\sigma}^2_l)$, تعریف کنیم، آنگاه $(M_k \leq M_l)$, $FIC(k) = \dfrac{(n - p_l)(\widehat{\sigma}^2_k - \widehat{\sigma}^2_l)}{(p_l - p_k)\widehat{\sigma}^2_l}$ قابل ذکر است که M_k و M_l برابر با $M_k : \boldsymbol{Y} = \mathbf{X}_k\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_k$ و $M_l : \boldsymbol{Y} = \mathbf{X}_l\boldsymbol{\beta}_l + \boldsymbol{\varepsilon}_l$ می باشند. علاوه بر این $\widehat{\boldsymbol{\beta}}_k, \widehat{\sigma}^2_k, \widehat{\boldsymbol{\beta}}_l$ و $\widehat{\sigma}^2_l$ برآوردگرهای درستنمایی ماکسیمم آنتروپی ماکسیمم متغیر پاسخ مدل رگرسیون M_k, M_l می باشند. بنابراین اگر $FIC(k) \leq FIC(l)$ باشد آنگاه مدل M_k بهینه تر خواهد بود.

می دانیم که هر چه خطای یک مدل کمتر یعنی مقدار پراکندگی آن کمتر باشد در نتیجه آن مدل بهینه تر خواهد بود. در اینجا معیارهایی برای انتخاب مدل رگرسیون مطرح می شود که بر اساس خطای مدل رگرسیون می باشد. آکائیک [13] و ساوا [14] با استفاده از اطلاع کولبک–

[11] Hotelling

[12] Mardia

[13] Akaike

[14] Sawa

لیبلر مقدار خطای مدل رگرسیون را به دست آورده اند و مدل رگرسیون بهینه را بر این مبنا انتخاب کرده اند. آکائیک در سال ۱۹۷۳ برابری زیر را برای دو برابر امیدریاضی اطلاع کولبک- لیبلر بین $f_k^*(\boldsymbol{y},\boldsymbol{\theta}_l)$ و $f_k^*(\boldsymbol{y},\widehat{\boldsymbol{\theta}_k})$ تعریف کرده است ($\nu_k = p_k + ۱, \nu_l = p_l + ۱$).

$$۲E_{\widehat{\boldsymbol{\theta}}_k}\{K[f_k^*(\boldsymbol{y},\boldsymbol{\theta}_l) : f_k^*(\boldsymbol{y},\widehat{\boldsymbol{\theta}}_k)]\} \approx -\frac{۲}{n}\log\frac{f_k^*(\boldsymbol{y},\widehat{\boldsymbol{\theta}}_k)}{f_k^*(\boldsymbol{y},\widehat{\boldsymbol{\theta}}_l)} + \frac{۲\nu_k}{n} - \frac{\nu_l}{n}. \qquad (۱.۷)$$

که رابطه (۱۰۷) بر اساس آنتروپی کولبک- لیبلر و اطلاع فیشر به دست می آید. آکائیک در رابطه (۱۰۷) n، ν_l و $f_k^*(\boldsymbol{y},\widehat{\boldsymbol{\theta}}_l)$ را ثابت در نظر گرفته و در نهایت خطای مدل را با فرمول $AIC(k) = -۲\log f_k^*(\boldsymbol{y},\widehat{\boldsymbol{\theta}_k}) + ۲\nu_k, \quad \nu_k = ۱,۲,\ldots$ نشان داد. به عنوان مثال برای مدل رگرسیون $M_k = N(\mathbf{X}_k\boldsymbol{\beta}_k, \sigma_k^۲\boldsymbol{I}_n), (\nu_k = p_k + ۱)$، خطای مدل آکائیک برابر با $AIC(k) = -۲\log f^*(\boldsymbol{y},\widehat{\boldsymbol{\theta}_k}) + ۲\nu_k = ۲H(f_k^*(\boldsymbol{y},\widehat{\boldsymbol{\beta}}_k,\widehat{\sigma}_k^۲)) + ۲(p_{k+۱}).$ است که به ازای مقادیر مختلف k خطای مدل رگرسیون به دست می آید که هر چه آنتروپی کمتر باشد (بیشترین اطلاع)خطای آکائیک مدل رگرسیون کمتر است در نتیجه مدل رگرسیون بهینه تر خواهد بود. ساوا (۱۹۷۸) برابری زیر را برای دو برابر امیدریاضی اطلاع کولبک–لیبلر بین $f(\boldsymbol{y},\boldsymbol{\mu},\omega)$ و $f(\boldsymbol{y},\tilde{\boldsymbol{\beta}},\tilde{\sigma^۲})$ بیان کرده است:

$$\begin{aligned}۲E_{\tilde{\boldsymbol{\beta}},\tilde{\sigma}^۲}\overline{K}\{f(\boldsymbol{y},\boldsymbol{\mu},\omega) : f(\boldsymbol{y},\tilde{\boldsymbol{\beta}},\tilde{\sigma}^۲)\} &= ۲\overline{H}(f(\boldsymbol{y},\boldsymbol{\mu},\omega)) + \log ۲\pi e + \log\sigma_۰^۲ + \frac{p_k+۱}{n}(\frac{\omega^۲}{\sigma_۰^۲})\\ &- \frac{۱}{n}(\frac{\omega^۲}{\sigma_۰^۲})^۲. \end{aligned} \qquad (۲.۷)$$

بنابراین با توجه به رابطه (۲۰۷) ساوا خطای مدل را به صورت

$$BIC(k) = -۲\log(f_k^*(\boldsymbol{y},\widehat{\boldsymbol{\beta}}_k,\widehat{\sigma}_k^۲)) + ۲(p_k+۱)(\frac{\omega^۲}{\widehat{\sigma}_k^۲}) - ۲(\frac{\omega^۲}{\widehat{\sigma}_k^۲})^۲,$$

تعریف کرد. به عنوان مثال برای مدل رگرسیون خطای $BIC(k)$ برابر با $BIC(k) = ۲H(f_k^*(\boldsymbol{y},\widehat{\boldsymbol{\beta}}_k,\widehat{\sigma}_k^۲)) + ۲(p_k+۱)(\frac{\widehat{w}^۲}{\widehat{\sigma}_k^۲}) - ۲(\frac{\widehat{w}^۲}{\widehat{\sigma}_k^۲})^۲$ است که در آن نسبت واریانس $\frac{\widehat{w}^۲}{\widehat{\sigma}_k^۲}$ نسبت به p_k نزولی می باشد. هر چه آنتروپی و نسبت واریانس کمتر باشند، خطای مدل رگرسیون کمتر است.

بحث و نتیجه‌گیری

با توجه به آنچه که مطرح کردیم برای برآورد ضرایب و واریانس مدل رگرسیون (برآورد آنتروپی نسبی مینیمم و برآورد آنتروپی نسبی مینیمم بیزی) و انتخاب مدل بهینه برای رگرسیون با استفاده از نظریه اطلاعات نیاز به اطلاع کولبک- لیبلر داریم و همچنین آزمون های آنتروپی نسبی مینیمم بیزی را می توان با کمک اطلاع جفریز بررسی کرد. قابل ذکر است که نتیجه ی آزمون فرضیه به روش آزمون کلاسیک همانند آزمون بیزی می باشد فقط در آزمون بیزی در آماره آزمون، برآورد آنتروپی نسبی مینیمم بیزی $\sigma^۲$ به کار رفته است. در نهایت می توان گفت که ارتباطی که بین رگرسیون و نظریه اطلاعات برقرار کردیم بر اساس اطلاع کولبک- لیبلر است.

مراجع

وحیدیان، ع. و بزرگ نیا، ا. (۱۳۷۲). نظریه کنترل و کنترل بهینه .

Akaike H. (1973), Information theory and an extension of maximum likelihood principle, 2nd International Symposium on Information Theory, 267-281.

Evren A. and Tuna E. (2012), Some applications of entropy- based statistics in linear regression analysis, Journal of Selcuk University Natural and Applied Science, 1, 39- 53.

Hartley R. V. L. (1928), Transmission of information, Bell System Technical Journal, 7, 535.

Jaynes E. T. (1968), On the rational of maximum entropy methods, Proceedings of Institute of Electrical and Electronics Engineers, 70, 939-952.

Kagan A. M., Linnik Y. V. and Rao C. R. (1973), Characterization Problems in Mathematical Statistics, Wiley, New York.

Kullback S. and Rosenblatt H. M. (1957), On analysis of multiple regression in k categories, Biometrika, 44, 67-83.

Kullback S. and Leibler R. A. (1951), On information and sufficiency, Annals of Mathematical Statistics, 22, 79-86.

Mardia K. V., Kent J. T. and Bibby J. M. (1979), Multivariate Analysis, Academic Press.

Nyquist H. (1924), Certain factors affecting telegragh speed, Bell System Technical Jourdan, 3, 126-147.

Sawa T. (1978), Information criteria for discriminating among alternative regression models, Journal of Econometrica, 46, 273-292.

Shannon C. E. (1948), mathematical theory of communication, Bell System Technical Journal, 27, 379-423.

Soofi E. S. (1996), Information theoretic regression methods, School of Business Administration University of Wisconsin-Milwaukee, 12, 1-67.