آزمونهای نیکوئی تابع پیوند

منوچهر خردمندنیا   مصطفی رستگاری [1]

چکیده

در خانواده الگوهای خطی تعمیم یافته، یکی از مراحل بررسی درستی تشخیص مدل، می تواند آزمون نیکوئی تابع پیوند باشد. در این مقاله، ضمن معرفی چند خانواده از توابع پیوند، مبانی نظری این آزمون را معرفی نموده آنرا برای الگوسازی تصادفات جاده‌ای ایران به کار می‌بریم.

۱ مقدمه

فرضیات و عناصر تشکیل دهنده یک الگوی خطی تعمیم یافته به شرح زیر است:

الف– فرض می‌شود y_1 تا y_N مستقلند و $E(y_i) = \mu_i$، بنابراین می‌توان نوشت:

$$y_i = \mu_i + \varepsilon_i \quad ; \quad E(\varepsilon_i) = 0$$

y_i را متغیر پاسخ، μ_i را قسمت سیستماتیک مدل و ε_i را باقیمانده می‌نامند.

ب– فرض می‌شود که تابع احتمال y_i را می‌توان به فرم

$$f(y_i;\theta_i,\phi) = exp\{\frac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + c(y_i,w_i,\phi)\}$$

نوشت، یعنی یکی از فرضیات اینست که تابع احتمالی y_i به شکل متعارف و حالتی منظم از رده نمایی است. $b(.)$ و $c(.)$ توابع معلومی هستند، θ_i پارامتر اصلی مورد علاقه است که آنرا پارامتر طبیعی می‌نامند. پارامتر طبیعی می‌تواند با i تغییرکند، ϕ پارامتر دیگری است که آزاد از i است و آنرا پارامتر مقیاس و بعضی پارامتر مزاحم می‌نامند. w_i ها اعداد حقیقی مثبت هستند که آنها را وزنهای قبلی می‌نامند.

ج– فرض می‌شود مکانیزمی که متغیرهای توضیحی معلوم $x_1, x_2, \ldots, x_p$ را به قسمت سیستماتیک مدل مرتبط می‌کند به فرم زیر است.

$$g(\mu_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$$

[1] منوچهر خردمندنیا و مصطفی رستگاری، گروه آمار– دانشکده علوم، دانشگاه اصفهان

که در آن $\eta_i = g(\mu_i)$ تابعی یکنوا و مشتق پذیر از μ_i است که آنرا تابع پیوند می‌نامند.

به عنوان مثالی از یک الگوی خطی تعمیم یافته، فرض کنید y_1 تا y_N مستقلند و $y_i \sim N(\mu_i, \frac{\sigma^2}{n_i})$. یعنی در واقع y_i میانگین یک نمونه تصادفی n_iتایی از $N(\mu_i, \sigma^2)$ می‌باشد. تابع چگالی احتمالی y_i عبارتست از:

$$\begin{aligned} f(y_i;\mu_i,\sigma^2) &= \frac{\sqrt{n_i}}{\sigma\sqrt{2\pi}} exp\{-\frac{(y_i-\mu_i)^2}{2\sigma^2/n_i}\} \\ &= exp\{\frac{y_i\mu_i-\mu_i^2/2}{\sigma^2/n_i} \\ &- \frac{1}{2}(\frac{n_i y_i^2}{\sigma^2}+ln(2\pi\sigma^2/n_i))\} \end{aligned}$$

بنابراین، در اینجا داریم: $\theta_i = \mu_i$، $\phi = \sigma^2$، $w_i = n_i$ و $b(\theta_i) = \frac{1}{2}\mu_i^2$. اکنون الگوی غیر خطی زیر را در نظر بگیرید

$$y_i = \sqrt{\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}} + \varepsilon_i$$

در اینجا داریم: $\mu_i = \sqrt{\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}}$ و یا به عبارت دیگر

$$\mu_i^2 = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$$

برای این الگو که الگوی توان دوم خطی نامیده می‌شود تابع پیوند $g(\mu_i) = \mu_i^2$ می‌باشد.

به عنوان مثالی دیگر، فرض کنید y_1 تا y_N مستقلند و y_i دارای توزیع پواسن با امید ریاضی μ_i است. تابع جرم احتمال y_i عبارتست از:

$$f(y_i;\mu_i) = \mu_i^{y_i} e^{-\mu_i}/y_i! = exp\{y_i ln\mu_i - \mu_i - lny_i!\}$$

بنابراین در اینجا داریم: $\theta_i = ln\mu_i$، $\phi = 1$، $w_i = 1$، $b(\theta_i) = \mu_i$ و $c(y_i, w_i, \phi) = -lny_i!$. اکنون الگوی غیر خطی

$$y_i = (\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip})^4 + \varepsilon_i$$

را در نظر بگیرید. در اینجا داریم:

$$\mu_i = (\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip})^4$$

و یا به عبارت دیگر

$$\sqrt{\sqrt{\mu_i}} = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$$

برای این الگو که الگوی ریشه چهارم خطی نامیده می‌شود تابع پیوند: $g(\mu_i) = \sqrt{\sqrt{\mu_i}}$ می‌باشد. به منظور سهولت مباحث نظری، در باقیمانده بخش حاضر، اندیس i را از مباحث خویش حذف می‌کنیم.

یکی از اهداف الگوسازی آماری در خانواده الگوهای خطی تعمیم یافته، تعیین تابع یکنوا و مشتق پذیر $g(\mu)$ از امید ریاضی پاسخ است بطوریکه به شکل خطی با متغیرهای توضیحی معلوم $x_1, x_2, \ldots, x_p$ مرتبط باشد. نلدر و ودربورن (۱۹۷۲)، تابع $g(\mu)$ را تابع پیوند نامیدند. برای الگوی رگرسیون معمولی با پاسخ نرمال، تابع پیوند همانی $g(\mu) = \mu$ مرسوم‌ترین تابع پیوند برای شروع است. لذا الگوی زیر را داریم که می‌توان آنرا الگوی همانی خطی نامید.

$$\mu = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \quad (1)$$

در تحلیل داده‌های شمارشی در جداول چند طرفه توافقی، تابع پیوند لگاریتم $g(\mu) = ln\mu$، تابع پیوند مناسبی است. به عبارت دیگر، الگوی زیر که الگوی لگاریتم خطی نامیده می‌شود در این حالت مناسب می‌باشد.

$$ln\mu = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \quad (2)$$

اهمیت تابع پیوند لگاریتم برای پاسخهای پواسن در جداول چند طرفه توافقی، به واسطه نقشی است که این تابع پیوند در تشخیص انواع استقلال ایفا می‌کند. در تحلیل پاسخهایی که به صورت «نسبت موفقیت» هستند، تابع پیوند لوجیت

$$g(\mu) = ln(\frac{\mu}{n-\mu}) = ln(\frac{\pi}{1-\pi}) = logit(\pi) \quad (3)$$

برای شروع مناسب و مرسوم است. به عبارت دیگر برای پاسخهای دوجمله‌ای الگوی زیر که آنرا می‌توان الگوی لوجیت خطی نامید، مناسب برای شروع است.

$$ln(\frac{\pi}{1-\pi}) = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \quad (4)$$

که در آن $\pi = \mu/n$ «احتمال موفقیت» است. اهمیت تابع پیوند لوجیت برای پاسخهای دو جمله‌ای، به این دلیل است که با به کاربردن این تابع پیوند، برآورد درست نمایی ماکزیمم π همواره در فاصله مجاز $(\circ, 1)$ قرار می‌گیرد. البته برای پاسخ دوجمله‌ای تابع "پیوندلگ‌لگ‌متمم" یعنی

$$g(\mu) = ln[-ln(1-\pi)] \qquad (۵)$$

و نیز تابع پیوند پروبیت

$$g(\mu) = \phi^{-1}(\pi) \qquad (۶)$$

نیز مناسب و مرسوم هستند. که در آن

$$Z \sim N(\circ, 1) \quad , \quad \pi = \phi(g(\mu)) = P(Z < g(\mu))$$

برای پاسخهای گاما، تابع پیوند وارون $g(\mu) = 1/\mu$ برای شروع مناسب و مرسوم است. به عبارت دیگر الگوی وارون خطی زیر مناسب است.

$$\frac{1}{\mu} = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \qquad (۷)$$

توابع پیوند همانی، لگاریتم، وارون و لوجیت، به ترتیب برای پاسخهای نرمال، پواسن، گاما و دوجمله‌ای، تابع پیوند طبیعی نامیده می‌شوند. دلیل این نامگذاری اینست که هر یک از این توزیع‌ها، عضوی از خانواده نمایی هستند واین توابع پیوند در واقع پارامتر طبیعی این توزیعها هستند.

۲ چند خانواده از توابع پیوند

در اغلب متون آماری مربوط به الگوسازی، یک تابع پیوند خاص، مناسب فرض می‌شود. بعضی از این توابع پیوند را در مقدمه ملاحظه کردیم. در این بخش چند خانواده از توابع پیوند را معرفی می‌کنیم. هر یک از این خانواده‌ها را برای ارزیابی تابع پیوند فرضی خاصی، می‌توان مورد استفاده قرار داد.

یک خانواده مخصوص و ساده از توابع پیوند، خانواده چند جمله‌ای توکی (۱۹۴۹) است

$$g(\mu;\gamma) = \mu + \gamma\mu^2 \qquad (۸)$$

در این خانواده به ازای $\gamma = \circ$ تابع پیوند همانی بدست می‌آید. خانواده مفید دیگری از توابع پیوند که برای پاسخهای مثبت نظیر پواسن و گاما به کار می‌آید، خانواده توانی توکی (۱۹۷۷) می‌باشد.

$$g(\mu;\gamma) = \begin{cases} \mu^\gamma & ; \quad \gamma \neq \circ \\ ln\mu & ; \quad \gamma = \circ \end{cases} \qquad (۹)$$

در این خانواده، به ازای $\gamma = \circ$، $\gamma = 1$، $\gamma = -1$ و $\gamma = \circ/5$ به ترتیب توابع پیوند لگاریتم، همانی، وارون و ریشه دوم به دست می‌آیند. توجه کنید که برای اعداد حقیقی دلخواه α و $\beta \neq \circ$، دو تابع پیوند $g(\mu;\gamma)$ و $\alpha + \beta g(\mu;\gamma)$ معادل یکدیگرند.

برای بحث مشابهی، نه مربوط به توابع پیوند، بلکه مربوط به تبدیلات، جمشیدیان (۱۹۹۲) را ملاحظه کنید. بخصوص در حالت خاص $\alpha = \circ$ و $\beta = 1/\gamma$ دو تابع پیوند μ^γ و μ^γ/γ معادل هستند. اما حد μ^γ/γ وقتی $\gamma \to \circ$ برابر $ln\mu$ است. به این ترتیب در خانواده توابع پیوند (۹)، تساوی $g(\mu;\circ) = ln\mu$ توجیه می‌شود. یک خانواده از توابع پیوند که استفاده از آن محدود به پاسخهای مثبت نیست، خانواده نمایی منلی (۱۹۷۵) می‌باشد.

$$g(\mu;\gamma) = \begin{cases} exp(\mu\gamma) & ; \quad \gamma \neq \circ \\ \mu & ; \quad \gamma = \circ \end{cases} \qquad (۱۰)$$

این خانواده بخصوص برای پاسخهای نرمال مفید است. در اینجا، حد $e^{\mu\gamma}/\gamma$ وقتی $\gamma \to \circ$ برابر μ است؛ لذا تساوی $g(\mu;\circ) = \mu$ توجیه می‌شود. خانواده دیگری از توابع پیوند که بخصوص برای پاسخهای دوجمله‌ای مفید است خانواده پرگیبون (۱۹۸۵) می‌باشد.

$$g(\pi;\gamma) = \begin{cases} \pi^\gamma + (1-\pi)^{-\gamma} & ; \quad \gamma \neq \circ \\ ln(\pi/(1-\pi)) & ; \quad \gamma = \circ \end{cases} \qquad (۱۱)$$

در اینجا نیز $\pi = \mu/n$ نشان دهنده «احتمال موفقیت» می‌باشد. در این خانواده، به ازای $\gamma = \circ$ تابع پیوند لوجیت

بدست می‌آید. در واقع حد $[\pi^{\gamma} + (1-\pi)^{-\gamma}]/\gamma$ وقتی $\gamma \to 0$ برابر $ln(\pi/(1-\pi))$ می‌باشد.

خانواده دیگری از توابع پیوند که برای پاسخهای دو جمله‌ای مفید است، خانواده توابع پیوند مک کولا و نلدر (۱۹۸۹) می‌باشد که عبارتست از:

$$g(\mu;\gamma) = ln[\frac{(1-\pi)^{-\gamma}-1}{\gamma}] \qquad (12)$$

این خانواده شامل تابع پیوند لوجیت (به ازای $\gamma = 1$) و تابع پیوند لگ لگ متمم (وقتی $\gamma \to 0$) می‌باشد.

۳ الگوسازی

فرض کنید $S(C,F)$ آماره منهای دو برابر لگاریتم نسبت درستنمایی برای مقایسه الگوی اشباع نشده M_C با الگوی اشباع شده M_F است. یعنی:

$$S(C,F) = -2ln(\ell_C/\ell_F)$$

که در آن، ℓ_C درستنمایی الگوی M_C و ℓ_F درستنمایی الگوی M_F است. اگر M_2 در M_1 آشیانه‌ای و $S(2,1)$ آماره منهای دو برابر لگاریتم نسبت درستنمایی برای مقایسه M_1 و M_2 باشد داریم:

$$S(2,1) = S(2,F) - S(1,F) = -2ln(\ell_2/\ell_1)$$

که ℓ_i درستنمایی الگوی M_i است. کمیت $S(C,F)$ را به شکل $S(C,F) = D(C,F)/\phi$ می‌توان نوشت. به طریق مشابه کمیت $S(2,1)$ را به شکل $S(2,1) = D(2,1)/\phi$ می‌توان نوشت.

نلدر و ودربورن (۱۹۷۲)، کمیت $D(C,F)$ را Deviance نامیدند. برگردان فارسی مناسبی برای این اصطلاح «انحراف درستنمایی» است که مقصود میزان انحراف درستنمایی الگوی M_C از درستنمایی الگوی M_F می‌باشد.

با توجه به اینکه M_2 در M_1 آشیانه‌ایست واضح است که $S(1,F) < S(2,F)$ و $P_1 > P_2$ که در آن P_1 و P_2 به ترتیب تعداد پارامترها (تعداد β ها)ی مدلهای M_1 و M_2 می‌باشند.

تحت فرض نیکوئی برازش M_C، متغیر تصادفی $S(C,F)$ بطور مجانبی دارای توزیع خی‌دو با $N - P_C$ درجه آزادی است که در آن P_C تعداد پارامترهای الگوی M_C است. بنابراین فرض نیکوئی برازش الگوی M_C در سطح α رد می‌شود هرگاه، مقدار مشاهده شده $S(C,F)$ بزرگتر از نقطه $100(1-\alpha)$ درصد توزیع خی‌دو با $N - P_C$ درجه آزادی باشد. برای پاسخهای پواسن، دوجمله‌ای و نمائی داریم $\phi = 1$. برای پاسخهای نرمال، گاما و وارون گوس $S(C,F)$ به پارامتر مزاحم ϕ نیز بستگی دارد. در این حالت به جای ϕ می‌توان برآوردگر سازگاری از آن را قرار داد. اکنون، فرض کنید فرض نیکوئی برازش هر دو الگوی M_1 و M_2 پذیرفته شده است. فرض معنی‌دار نبودن اثرات اضافی بین این دو الگو در سطح α رد می‌شود هرگاه، مقدار مشاهده شدهٔ $S(2,1)$، بزرگتر از نقطه $100(1-\alpha)$ درصد توزیع خی‌دو با $P_1 - P_2$ درجه آزادی باشد. در اینجا نیز وقتی توزیع متغیر پاسخ به ϕ بستگی دارد، برآوردگر سازگاری از ϕ را می‌توان بجای آن قرار داد. ولی در این حالت، برای آزمون فرض معنی‌دار نبودن اثرات اضافی بین M_1 و M_2 آنچه که در متون آماری مرسوم‌تر است، مقایسه مقدار مشاهده شده $D(2,1)/\tilde{\phi}(P_1 - P_2)$ با عدد بحرانی مناسبی از جدول توزیع F می‌باشد؛ که در آن $\tilde{\phi}$ برآوردی سازگار برای ϕ می‌باشد.

۴ آزمون نیکویی تابع پیوند

فرض کنید الگویی با تابع پیوند $g(\mu;\gamma_0)$ را به عنوان الگوی نهایی تعیین کرده‌ایم که در آن، $g(\mu;\gamma_0)$ عضوی از خانواده $g(\mu;\gamma)$ است. می‌خواهیم بررسی کنیم که آیا در همین خانواده، جایگزینی تابع پیوند جدید $g(\mu;\gamma_*)$، الگوی نهایی را بطور معنی‌داری بهتر می‌کند یا خیر؛ که در آن $g(\mu;\gamma_*)$ عضوی از همان خانواده $g(\mu;\gamma)$ است و γ_* در نزدیکی γ_0 است. الگویی که با فرض صفر رقابت می‌کند را به صورت زیر می‌توان نوشت:

$$g(\mu;\gamma_*) = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \qquad (13)$$

تقریب مرتبه اول بسط تیلور $g(\mu;\gamma_*)$ حول $\gamma_\circ$ عبارتست از:

$$g(\mu;\gamma_*) \cong g(\mu;\gamma_\circ) + (\gamma_* - \gamma_\circ)g'(\mu;\gamma_\circ) \qquad (\text{۱۴})$$

که در آن

$$g'(\mu;\gamma_\circ) = \frac{d}{d\gamma}g(\mu;\gamma)|_{\gamma=\gamma_\circ}$$

با جایگذاری از (۱۴) در (۱۳) داریم:

$$g(\mu;\gamma_\circ) =$$

$$\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + (\gamma_\circ - \gamma_*)z \qquad (\text{۱۵})$$

که در آن $z = g'(\mu;\gamma_\circ)$. برآورد z تحت فرض صفر برابر است با $\hat{z} = g'(\hat{\mu};\gamma_\circ)$، که در آن $\hat{\mu}$ برآورد ML پارامتر μ تحت فرض صفراست. اگر در (۱۵) به جای z، برآورد آن $\hat{z}$ را قراردهیم الگوی زیر حاصل می‌شود:

$$g(\mu;\gamma_\circ) = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \delta\hat{z} \qquad (\text{۱۶})$$

که در آن $\delta = \gamma_\circ - \gamma_*$. با در نظر گرفتن $\hat{z}$ به عنوان یک متغیر توضیحی جدید، پارامتر δ را می‌توان برآورد نمود و سپس فرض $\delta = \circ$ را آزمود. رد شدن فرض $\delta = \circ$ بدین معنی است که یک γ_* وجود دارد، بطوریکه $g(\mu;\gamma_*)$ بطور معنی‌داری بهتر از $g(\mu;\gamma_\circ)$ است. بنابراین، در صورت رد شدن فرض $\delta = \circ$، می‌بایست γ_* را با استفاده از تساوی $\hat{\gamma}_* = \gamma_\circ - \hat{\delta}$ برآورد نمود که در آن $\hat{\delta}$ برآورد ML پارامتر δ می‌باشد. سپس، باید الگوسازی را با تابع پیوند $g(\mu;\gamma_*)$ تجدید نمود. به عنوان مثال در خانواده توابع پیوند توانی توکی، برای فرضهای صفر $g(\mu;1) = \mu$، $g(\mu;\circ) = ln\mu$ و $g(\mu;-1) = \frac{1}{\mu}$ به ترتیب داریم: $z = \mu ln\mu$، $z = ln^2(\mu)$ و $z = (ln\mu)/\mu$. در خانواده توابع پیوند نمایی منلی، برای فرضهای صفر $g(\mu;1) = e^\mu$، $g(\mu;\circ) = \mu$ و $g(\mu;-1) = e^{-\mu}$ به ترتیب داریم: $z = \mu e^\mu$، $z = \mu^2$ و $z = \mu e^{-\mu}$.

۵ الگوسازی تصادفات جاده‌ای ایران

در این بخش، به منظور ارائه مثالی از آزمون نیکویی تابع پیوند، تعداد تصادفات جاده‌ای ایران را مورد الگوسازی قرار می‌دهیم. تعداد سالانه تصادفات جاده‌ای ایران در فاصله سالهای ۱۳۵۲ ($t =$ ۱) لغایت ۱۳۶۴ ($t =$ ۱۳) بر حسب صد نفر، در جدول(۱) و نمودار(۱) ملاحظه می‌شود. در نمودار (۱) سری زمانی مشاهده شده با علامت (o) مشخص شده است. این آمار مربوط به تصادفاتی است که منجر به حداقل یک فوت و یا یک مجروح انتقالی به بیمارستان شده است. داده‌ها در کتاب آیتی (۱۳۷۱) موجود است.

در اینجا، معقول است فرض کنیم که تعداد تصادفات در سال t تحققی از $Y_t \sim poisson(\mu_t)$ است و Y_tها مستقلند. با توجه به نمودار(۱) الگوی همانی خطی $\mu_t = \beta_\circ + \beta_1 t$، به نظر معقول می‌رسد. با استفاده از نرم‌افزار GLIM برآورد ML پارامترهای $\beta_\circ$ و β_1 به ترتیب برابر ۶۰/۹۴ و ۸/۵۸ به دست می‌آیند؛ انحراف معیارهای متناظر با این برآوردها به ترتیب برابر ۵/۵۴ و ۰/۷۹ می‌باشند. با توجه به اینکه برآوردگر ML پارامترها مجانباً نرمال است، فاصله (۱/۹۶(۵/۵۴) ± ۶۰/۹۴) یک فاصله اطمینان ۹۵ درصد برای $\beta_\circ$ است و فاصله (۱/۹۶(۰/۷۹) ± ۸/۵۸) یک فاصله اطمینان ۹۵ درصد برای β_1 می‌باشد. بنابراین، هریک از فرضهای صفر $\beta_\circ = \circ$ و $\beta_1 = \circ$ در سطح ۰/۰۵ رد می‌شوند. مدل برازش شده را به صورت زیر می‌توان نوشت:

$$\hat{\mu}_t = \underset{(\text{۵/۵۴})}{\text{۶۰/۹۴}} + \underset{(\text{۰/۷۹})}{\text{۸/۸۵}}\; t \qquad (\text{۱۷})$$

مقدار $-2ln\lambda$ و تعداد درجات آزادی متناظر با فرض نیکویی برازش این الگو به ترتیب برابر ۱۵/۹۷۵ و ۱۱ به دست می‌آیند. با توجه به اینکه $\chi^2_{\circ/\circ5}(11) =$ ۱۹/۶۷۵ لذا فرض نیکویی برازش این الگو در سطح ۰/۰۵ پذیرفته می‌شود. با افزودن متغیر توضیحی t^2 الگوی زیر به دست می‌آید.

$$\hat{\mu}_t = \underset{(\text{۸/۷۹})}{\text{۵۵/۱۱}} + \underset{(\text{۳/۲۲})}{\text{۱۱/۲۷}}\; t - \underset{(\text{۰/۲۴})}{\text{۰/۲۱}}\; t^2 \qquad (\text{۱۸})$$

اعدادی که در پرانتز و در زیر برآورد پارامترها نوشته شده، انحراف معیار متناظر آن برآورد است. با توجه به

اینکه فاصله $(-\circ/۲۱ \pm ۱/۹٦(\circ/۲۴))$ شامل صفر است، فرض صفر بودن ضریب $t^۲$ در سطح $\circ/\circ۵$ پذیرفته می‌شود. بنابراین، $t^۲$ حضور معنی داری ندارد. طریق دیگری که برای مقایسه دو الگوی فوق وجود دارد، این است که بررسی کنیم که آیا با افزودن $t^۲$ به مدل (۱۷)، مقدار $-۲ln\lambda$ بطور معنی‌داری کاهش می‌یابد یا خیر؟ مقدار $-۲ln\lambda$ متناظر با الگوی (۱۸) برابر ۱۵/۱۸۹ است. بنابراین مقدار کاهش یافته برابر $۱۵/۹۷۵ - ۱۵/۱۸۹ = \circ/۷۸٦$ می‌باشد و تعداد درجات آزادی کاهش یافته برابر ۱=۱۰−۱۱ است. با توجه به اینکه $\chi^۲_{\circ/\circ۵}(۱) = ۳/۸۴ > \circ/۷۸٦$؛ لذا فرض معنی‌دار نبودن اثر افزودن $t^۲$ به الگوی (۱۷) پذیرفته می‌شود. با بررسی متغیرهای توضیحی دیگری نظیر $\sqrt{t}$، $ln(t)$، $t^۳$ و ... مدل (۱۷)، بهبود معنی‌داری نمی‌یابد.

اکنون، به نظر می‌رسد جهت دیگری که برای بررسی امکان بهبود مدل وجود دارد بررسی توابع پیوند دیگر است. برای پاسخهای پواسن، خانواده توابع پیوند توانی (۹) مناسب است.

در این خانواده برای حالت $\gamma = \circ$ الگوی لگاریتم خطی $ln\mu_t = \beta_\circ + \beta_۱ t$ را داریم. مدل برازش شده لگاریتم خطی به صورت زیر است:

$$ln\hat{\mu}_t = \underset{(\circ/\circ۷)}{۴/۲۷} + \underset{(\circ/\circ\circ۷)}{\circ/\circ٦}\ t \qquad (۱۹)$$

برای این الگو داریم $-۲ln\lambda = ۲\circ/۲۱$ و $df = ۱۱$. با توجه به اینکه $\chi^۲_{\circ/\circ۵}(۱۱) = ۱۹/٦۷۵$ و $\chi^۲_{\circ/\circ۴}(۱۱) = ۲\circ/۴۱۲$، فرض نیکویی برازش الگوی لگاریتم خطی در سطح $\circ/\circ۵$ رد می‌شود ولی در سطح $\circ/\circ۴$ پذیرفته می‌شود. در مورد این الگو نیز هریک از فرضهای صفر $\beta_\circ = \circ$ و $\beta_۱ = \circ$ در سطح $\circ/\circ۵$ رد می‌شوند. برای حالت $\gamma \neq \circ$ الگوی $\mu_t^\gamma = \beta_\circ + \beta_۱ t$ را داریم. اکنون مسأله، یافتن بهترین مقدار $\gamma \neq \circ$ می‌باشد. در جدول (۲) و نمودار(۲) مقدار $-۲ln\lambda$ متناظر با مقادیر مختلف γ ملاحظه می‌شوند. خط افقی داخل نمودار(۲) نشان دهنده عدد بحرانی $\chi^۲_{\circ/\circ۵}(۱۱) = ۱۹/٦۷۵$ می‌باشد. ملاحظه می‌شود که برای $\gamma \in (a, b)$ فرض نیکویی برازش الگو در سطح $\circ/\circ۵$ پذیرفته می‌شود. مقادیر a و b روی محور افقی نمودار(۲) مشخص شده‌اند. اما پذیرفته‌شدن فرض نیکویی برازش یک الگو، فقط یکی از مراحل الگوسازی است. با استفاده از آزمون نیکویی تابع پیوند، مناسبت هریک از γها را می‌توان مورد بررسی قرار داد.

در این بخش آزمون نیکویی تابع پیوند را برای سه الگوی ریشه چهارم خطی ($\gamma = \circ/۲۵$)، الگوی توان دوم خطی ($\gamma = ۲$) و الگوی لگاریتم خطی ($\gamma = \circ$) به کار می‌بریم. الگوی ریشه چهارم خطی برازش شده بصورت زیر است:

$$\sqrt{\sqrt{\hat{\mu}_t}} = \underset{(\circ/\circ۴۷)}{۲/۸۸} + \underset{(\circ/\circ\circ٦)}{\circ/\circ۵۹}\ t \qquad (۲\circ)$$

با توجه به اینکه $a < \circ/۲۵ < b$ فرض نیکویی برازش این الگو در سطح $\circ/\circ۵$ پذیرفته می‌شود. با توجه به برآورد پارامترها و انحراف معیار متناظر آنها هریک از فرضهای صفر $\beta_\circ = \circ$ و $\beta_۱ = \circ$ در سطح $\circ/\circ۵$ رد می‌شوند. لذا به نظر می‌رسد که الگو قابل قبول باشد. اکنون نیکویی تابع پیوند $\sqrt{\sqrt{\mu_t}}$ را مورد آزمون قرار می‌دهیم؛ در اینجا داریم: $\hat{z}_t = \hat{\mu}_t^{\circ/۲۵} ln\hat{\mu}$ که در آن $\hat{\mu}_t = (۲/۸۸ + \circ/\circ۵۹t)^۴$ با افزودن $\hat{z}_t$ به الگوی (۲۰) الگوی برازش شده زیر بدست می‌آید:

$$\sqrt{\sqrt{\hat{\mu}_t}} = \underset{(۹/۸۷)}{۲۳/۴۷} + \underset{(\circ/۴۲)}{\circ/۹۴}\ t - \underset{(\circ/۸۱)}{۱/٦۷}\ \hat{z}_t \qquad (۲۱)$$

با توجه به اینکه فاصله $(-۱/٦۷ \pm ۱/۹٦(\circ/۸۱))$ شامل عدد صفر نیست لذا فرض $\delta = \circ$ در سطح $\circ/\circ۵$ رد می‌شود. بنابراین فرض نیکویی تابع پیوند $\sqrt{\sqrt{\mu_t}}$ در سطح $\circ/\circ۵$ رد می‌شود. برآورد ML پارامتر γ_* عبارتست از:

$$\hat{\gamma}_* = \gamma_\circ - \hat{\delta} = \circ/۲۵ - (-۱/٦۷) = ۱/۹۲$$

نتیجه می‌گیریم که تابع پیوند ریشه چهارم نامناسب است و حدس می‌زنیم که مقدار مناسب γ در نزدیکی ۱/۹۲ باشد. توجه کنید که الگوی ریشه چهارم خطی الگویی است که در سطح $\circ/\circ۵$ پذیرفته شده ولی براساس آزمون نیکویی تابع پیوند این الگو از لیست الگوهای قابل قبول حذف شده

است. با توجه به برآورد $\hat{\gamma}_* = ۱/۹۲$ به الگوی توان دوم خطی ($\gamma = ۲$) علاقمند می‌شویم. الگوی برازش شده توان دوم خطی به صورت زیر است:

$$\hat{\mu}_t^{۲} = \underset{(۸۹۹/۶)}{۲۳۳۱} + \underset{(۱۷۷/۶)}{۱۸۸۹}\ t \qquad (۲۲)$$

برای این الگو داریم: $\hat{z}_t = \hat{\mu}^{۲} ln\hat{\mu}$ که در آن $\hat{\mu}_t = \sqrt{۲۳۳۱ + ۱۸۸۹t}$، با افزودن $\hat{z}_t$ به الگوی (۲۲) الگوی برازش شده زیر به دست می آید:

$$\hat{\mu}_t^{۲} = \underset{(۵۳۹۰)}{-۴۸۰۶} - \underset{(۷۴۷۲)}{۷۸۱۱}\ t + \underset{(۰/۷۶۰۴)}{۰/۹۹}\ \hat{z}_t \qquad (۲۳)$$

ملاحظه می‌شود که فرض $\delta = ۰$ در سطح ۰/۰۵ پذیرفته می‌شود. بنابراین فرض نیکویی تابع پیوند μ_t را در سطح ۰/۰۵ می‌پذیریم. در نمودار(۱) مقادیر برازش شده الگوی توان دوم خطی با علامت ($\times$) مشخص شده و مقادیر برازش شده الگوی همانی خطی بصورت خطی مستقیم ملاحظه می‌شود. به عنوان مثالی دیگر از آزمون نیکویی تابع پیوند، فرض نیکویی تابع پیوند لگاریتم را بررسی می‌کنیم. مدل لگاریتم خطی برازش شده (۱۹) را در نظر بگیرید. برای آزمون فرض نیکویی تابع پیوند $ln\mu_t$ داریم $\hat{\mu}_t = exp\{۴/۲۷ + ۰/۰۶t\}$؛ که در آن $\hat{z}_t = (ln\hat{\mu}_t)^{۲}$ با افزودن $\hat{z}_t$ به الگوی (۱۹) الگوی برازش شده زیر بدست می آید.

$$ln\hat{\mu}_t = \underset{(۷/۶۸)}{۲۳/۱۴} + \underset{(۰/۲۹)}{۰/۷۸}\ t - \underset{(۰/۴۳)}{۱/۰۵}\ \hat{z}_t \qquad (۲۴)$$

مقدار احتمال برای آزمون فرض $H_۰ : \delta = ۰$ در مقابل $H_۱ : \delta \neq ۰$ برابر است با:

$$P.V = ۲P(z > \frac{۱/۰۵}{۰/۴۳}) = ۰/۰۱۴ \qquad (۲۵)$$

بنابراین فرض نیکویی تابع پیوند لگاریتم در سطح ۰/۰۴ رد می‌شود. این، در حالی است که فرض نیکویی برازش الگوی لگاریتم خطی (۱۹) در سطح ۰/۰۴ پذیرفته می‌شود. ملاحظه می‌شود که آزمون نیکویی تابع پیوند، به عنوان یکی از مراحل بررسی درستی تشخیص مدل، می‌تواند بعضی از الگوهای نادرستی که از فیلتر آزمون نیکویی برازش با موفقیت عبور کرده‌اند را به دام اندازد. با این حال هنوز هم، نه فقط یک مدل، بلکه چند مدل وجود دارند که از هر دو فیلتر آزمون نیکویی برازش و آزمون نیکویی تابع پیوند با موفقیت عبور کرده‌اند.

در اینصورت، بالاخره، مدل نهایی کدام است؟ در پاسخ به این سؤال باید گفت که آزمون نیکویی تابع پیوند فقط یکی از ابزارهای درستی تشخیص مدل است. ابزارهای دیگری، نظیر بررسی باقیمانده‌ها، وجود دارند که مورد بحث مقاله حاضر نیستند؛ آن ابزارها نیز به نوبه خود می‌توانند بعضی الگوهای نادرست را به دام اندازند. با وجود همه این ابزارها، باز هم چند مدل و نه فقط یک مدل باقی می‌ماند در اینجا، این دیگر مدلساز است که باید تصمیم نهایی را بگیرد و یکی از چند مدل را انتخاب کند. این انتخاب می‌تواند بر اساس معیار سادگی نسبی مدل باشد. یک روش محافظه‌کارانه این است که دو الگویی که دارای رفتار قرین و ناهمجهت هستند را انتخاب و نتایج را بر اساس آن دو الگو ارائه داد. به عنوان مثال، فرض کنید الگوهای $M_۱$ و $M_۲$ به ترتیب کمترین و بیشترین مقدار را برای متغیر پاسخ پیش بینی می‌کنند. در اینصورت دو مقداری که توسط این دو الگو پیش بینی می‌شوند، می‌توانند به عنوان دو انتهای یک دامنه پیش بینی برای پاسخ تلقی گردند.

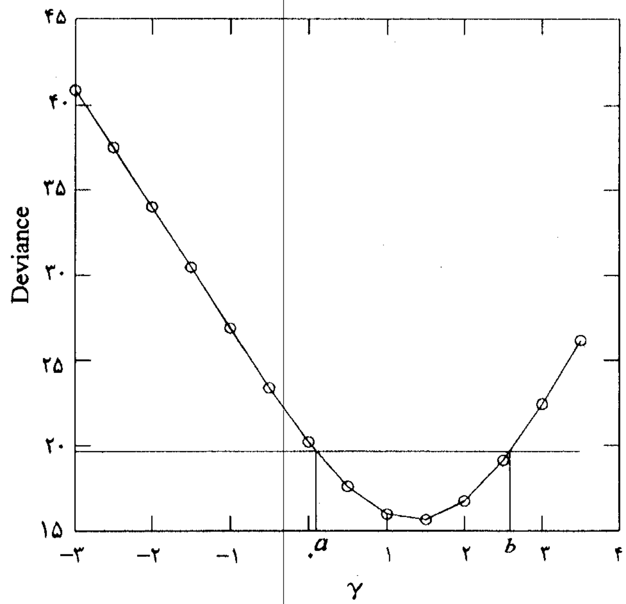
جدول ۱ – تعداد سالانه تصادفات جاده‌ای ایران در فاصله سالهای ۱۳۵۲ لغایت ۱۳۶۴

سال	۱۳۶۴	۱۳۶۳	۱۳۶۲	۱۳۶۱	۱۳۶۰	۱۳۵۹	۱۳۵۸	۱۳۵۷	۱۳۵۶	۱۳۵۵	۱۳۵۴	۱۳۵۳	۱۳۵۲
تعداد	۱۶۶	۱۵۸	۱۵۷	۱۳۱	۱۴۲	۱۴۷	۱۴۶	۱۰۱	۱۱۵	۸۶	۷۸	۶۸	۷۸

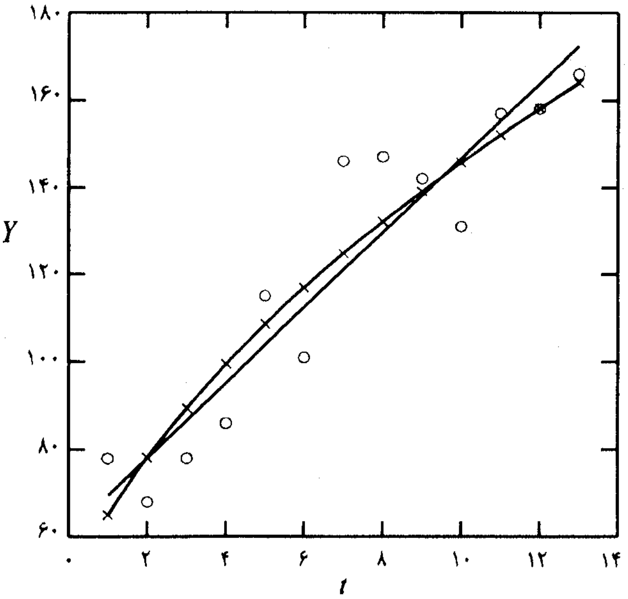
جدول ۲

γ	۰	−۰/۵	−۱	−۱/۵	−۲	−۲/۵	−۳
$-2ln\lambda$	۲۰/۲۱	۲۳/۳۹	۲۶/۸۶	۳۰/۴۴	۳۴/۰۰	۳۷/۵۰	۴۰/۸۷
γ	۳/۵	۳	۲/۵	۲	۱/۵	۱	۰/۵
$-2ln\lambda$	۲۶/۱۶	۲۲/۴۲	۱۹/۱۶	۱۶/۷۷	۱۵/۶۵	۱۵/۹۵	۱۷/۶۱

نمودار(۲)-مقادیر مشاهده شده انحراف درستنمائی در مقابل مقادیر مختلف γ



نمودار (۱)-مشاهدات (o) توان دوم خطی (×) و همانی خطی (خط مستقیم)



مراجع

[۱] آیتی، اسماعیل (۱۳۷۱) تصادفات جاده‌ای ایران، دانشگاه فردوسی مشهد.

[2] Agresti, A. (1990) *Categorical Data Analysis,* John Wiley and Sons.

[3] Baker, R. J. and Nelder, J. A. (1978) *The GLIM System, Generalized Linear Interactive Modelling.* Numerical Algorithms Group, Oxford.

[4] Jamshidian, M. (1992) Graphical Data analysis in Linear Regression, Proceedings of the first Iranian Statistics Conference, Isfahan University of Technology.

[5] Manly, B. F. J. (1975) Exponentioal Data Transformatio. *Statistician,* 25, 37-42.

[6] McCullagh, P. and Nelder J. A. (1989) *Generalized Linear Models.* 2nd ed. Chapman and Hall, London.

[7] Nelder, J. A. and Wedderburn, R. W. M. (1972) Generalized Linear Models. *J. R. Statist. Soc.* A, 135, 370-84.

[8] Pregibon, D. (1980) Goodness of Link Tests for Generalized Linear Models. *App. Statist.*, 29, 15-24.

[9] Pregibon, D. (1985) Link Tests, Encyclopedia of Statistical Sciences, Vol. 5.

[10] Tukey, J. W. (1949) One Degree of Freedom for Non-Additivity, *Biometrics*, 5, 232-42.

[11] Tukey, J. W. (1977) *Exploratory Data Analysis*. Addison-Wesley.