

صدکهای رگرسیونی با استفاده از تابع زیان مربع خطای نامتقارن

احمدرضا جمشیدیان^۱ سروش علیرادی^۱

چکیده

خروجی اولیه در تحلیل رگرسیونی معمولی برآورده تابع رگرسیونی است که نقاط وسط انبوه نقاط را نمایش می‌دهد. حال اگر علاقه مند به شرح نقاط پایینی یا بالایی به همان خوبی نقاط میانی باشیم، چگونه عمل می‌کنیم؟ بدین منظور براساس تعیین صدکهای نمونه‌ای به رگرسیون خطی، صدکهای رگرسیونی را تعریف می‌کنیم. در پایان اطلاعات جمع آوری شده از بیماران دیابتی پیمارستان امین اصفهان را برای مطالب ذکر شده به عنوان نمونه استفاده کردایم.

۱ مقدمه

افرون^۲ (۱۹۹۱) [۳] در این خصوص تحقیقاتی را انجام داده که مقاله حاضر برگرفته از مقاله اوست. در اینجا سعی بر این است که روش مفیدی را برای برآورد صدکهای رگرسیونی شرح دهیم. برای مثال ۲۵ امین یا ۷۵ امین صدک شرطی γ به عنوان تابعی از X برآورده شوند.

داده‌هایی که برای انجام یک تحلیل رگرسیونی به کارمی‌رونده فرم انبوه از نقاط در فضای اقلیدسی به صورت زیر هستند:

$$(x_i, y_i) \quad i = 1, 2, 3, \dots, n$$

که x یک بردار $p \times 1$ از متغیرهای مستقل و y یک مقدار حقیقی از متغیر پاسخ است. خروجی اولیه یک تحلیل معمولی برآورده تابع رگرسیونی است که نقاط وسط انبوه، نقاط را نمایش می‌دهند. اگر علاقه مند به شرح نقاط پایینی یا بالایی به همان خوبی نقاط میانی باشیم، چگونه عمل می‌کنیم؟

شکل (۱-۱) یک رگرسیون ساده را نشان می‌دهد. «ساده» بدین مفهوم که بردار x تک بعدی است و داده‌ها را می‌توان به وسیله نمودار پراکندگی، به صورت زوجهای (x_i, y_i) نشان داد. این شکل مربوط به یک آزمایش کلینیکی در استفاده از

^۱- احمد رضا جمشیدیان، سروش علیرادی، گروه آمار دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان
^۲- B. Efron-

مورد نظر است، برآورده شده‌اند. صدکهای رگرسیونی اطلاعات بیشتری نسبت به خط حداقل مربعات معمولی در اختیار ما قرار می‌دهند. در این آزمایش اگر بیماری α درصد از دارویی را که قبول کرده، مصرف کند، با تمرکز β درصد پاسخ، γ امین تا δ امین صدک برآورده می‌شود تا یک کاهش کلسترول خون مابین ϵ تا ζ واحد در مجموعه کل کلسترولهای شمارش شده باشد.

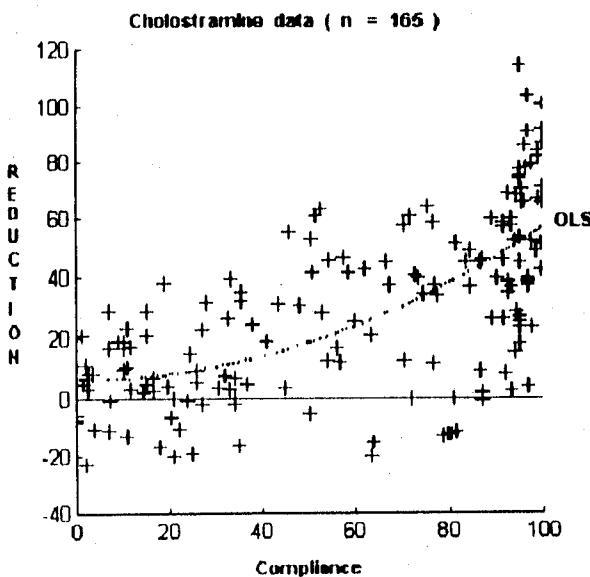
روش متداول برای نشان دادن صدکهای رگرسیونی بدین صورت است. فرض کنید میزان خطا در مدل رگرسیون همگن و نرمال (یعنی $(\epsilon_i \sim N(0, \sigma^2))$ برای $i = 1, 2, 3, \dots, n$) باشد، برآورد σ^2 را بصورت $\hat{\sigma}^2$ نشان می‌دهیم، آنگاه برآورد α γ امین صدک رگرسیونی به صورت:

$$\hat{\mu}(x) + \hat{\sigma} Z^{(\alpha)} \quad (1-1)$$

به دست می‌آید [۹] که (x) برآورد رگرسیون حداقل مربعات معمولی و $Z^{(\alpha)}$ هم α امین صدک از تابع چگالی نرمال استاندارد $Z^{(\alpha)} = \Phi^{-1}(\alpha)$ است. اگر تمام این خطوط را برای α های متفاوت در یک صفحه محورهای مختصات رسم کنیم تمام آنها موازی و فاصله آنها $\hat{\sigma} Z^{(\alpha)}$ است.

ایده اصلی صدکهای رگرسیونی بر اساس ابداعات کانکر و باست [۱۹۷۸] [۶] می‌باشد. آنها از قدر مطلق زیان نامتقارن θ بیشتر از مربع زیان نامتقارن α استفاده کرده، در این بین چندکهای رگرسیونی θ را تعریف کردند. پاول و نیوی [۱۰] [۱۹۸۷] کاری را که توسط آیگنر، آمیما و پویربر [۱۱] [۱۹۷۶] انجام شده بود، ادامه دادند و از مربع

داروی کلسترولین θ و تأثیر آن بر پایین آوردن کلسترول خون θ است. بنابراین y میزان کاهش خون و x درصد بسته‌های کلسترولین مصرفی است که بیمار θ ممکن قبول کرده است. منحنی حداقل مربعات معمولی برآزandه شده، یک مدل رگرسیونی درجه دو به صورت $y_i = b_0 + b_1 x_i + b_2 x_i^2 + e_i$ است.^۷



شکل ۱-۱: این شکل مربوط به آزمایش کلینیکی در استفاده از داروی کلسترولین و تأثیر آن بر پایین آوردن کلسترول خون، همراه با منحنی حداقل مربعات معمولی برآزandه شده بصورت مدل رگرسیونی درجه دو، ممکن قبول کرده است. $y_i = b_0 + b_1 x_i + b_2 x_i^2 + e_i$

شکل (۱-۲) صدکهای رگرسیونی این داده‌ها را نشان می‌دهد که به وسیله روش حداقل مربعات نامتقارن، که در اینجا

۱- این اطلاعات مربوط به بیماران تحت کنترل بخش استانفورد می‌باشد که در Lipid Research Clinics Program از دانشکده داروسازی استانفورد جمع آوری شده است.

Cholostyramine-۲	R. Koenker and G. Bassett-۸
Blood Cholesterol-۴	Asymmetric Absolute Loss-۹
Reduction-۵	Asymmetric Squared Loss-۱۰
Compliance-۶	Regression Quantiles-۱۱
-۷	Powell and Newey-۱۲
	Aigner , Amemiya and poirier-۱۳

خطی تعمیم داد و آن را صدکهای رگرسیونی نامید. این صدکها مفسرها مفید و خوبی برای مجموعه داده‌های رگرسیونی می‌باشند و سعی داریم که الگوریتم ساده‌ای را برای همه صدکهای رگرسیونی مربع خطای نامتقارن برای یک مجموعه از داده‌ها ارایه دهیم. بیشتر کار ما در ادامه مقاله پاول و نیوی [۸] است.

خطای نامتقارن آنچنان که ما در اینجا از آن استفاده خواهیم کرد، استفاده کردند و نتایج منحنی را انتظارک^{۱۲} نامیدند، آنها هرچند در ابتدا در قسمت «صدکهای رگرسیونی» قصد داشتند که فرم‌های متنوعی را از ایده اصلی به کار گیرند ولی در ادامه بیشتر به منحنی‌های حساب شده بوسیله حداقل مربعات نامتقارن پرداختند. در عمل اکثر این محققین به دنبال استوارسازی^{۱۵} [۵] پارامترهای رگرسیونی بوده‌اند.

۲ صدکهای رگرسیونی

برای تعریف صدکهای رگرسیونی مدل رگرسیون خطی زیر را در نظر می‌گیریم:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \quad (1-2)$$

که فرم برداری آن به صورت زیر است:

$$y_i = x'_i \beta + \varepsilon_i \quad x_{i1} = 1 \quad \& \quad \varepsilon_i \sim (0, \sigma^2) \quad i = 1, 2, \dots, n \quad (2-2)$$

و فرم ماتریس مدل هم به صورت زیر می‌باشد:

$$Y = X\beta + \varepsilon \quad (3-2)$$

که $(\varepsilon_1, \dots, \varepsilon_n) = \varepsilon$ ، بردار خطایی است دارای تابع توزیع F با میانگین صفر و واریانس ثابت σ^2 ، $\beta = (\beta_1, \dots, \beta_n)$ بردار پارامترهای مجهول، X یک ماتریس $n \times p$ معلوم بارتیه کامل سنتوتی و $(y_1, \dots, y_n) = Y$ بردار مشاهدات است.

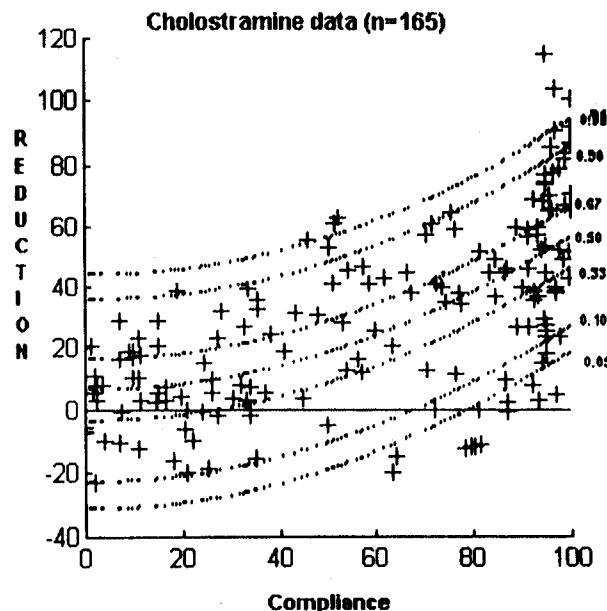
برای برآورد صدکهای رگرسیونی پارامتر β ، ابتدا تابع زیان

مربع خطای نامتقارن را برای هر β به فرم:

$$r_w(\lambda) = \begin{cases} \lambda^2 & \lambda \leq 0 \\ w\lambda^2 & \lambda > 0 \end{cases} \quad (4-2)$$

تعریف کرده و سپس معادله:

$$S_w(\beta) = \sum_{i=1}^n r_w(y_i - x'_i \beta) \quad (5-2)$$



شکل ۱-۲: این شکل مربوط به صدکهای رگرسیونی متداول برای آزمایش کلینیکی در استفاده از داروی کلسترول خون است و برای بروپاپین آوردن کلسترول خون است و برای $\alpha = 0.05, 0.10, 0.22, 0.50, 0.67, 0.90, 0.95$

صدکهای رگرسیونی آن رسم شده است.

در اینجا صدکهای رگرسیونی که اولین بار توسط افرون (۱۹۹۱)^{۱۳} ارایه شد جالبترند. وی صدکهای نمونه‌ای را که در توصیف اطلاعات نمونه‌ای مفیدند، براساس چندکهای رگرسیونی تعریف شده توسط کانکرو باست [۶] به رگرسیون

Expectile—^{۱۲}
Robustify—^{۱۵}

$\hat{\mathcal{L}}^{\alpha}$ نیز نشان داد و این نامگذاری براساس نسبت مشاهدات (x_i, y_i) که پایین این صفحات واقع شده‌اند، می‌باشد.

برای محاسبه صدکهای رگرسیونی، اگر فرض کنیم b یکی از مقادیر مورد نظر برای معادله $(2-2)$ باشد، آنگاه بردار باقی مانده‌ها (b) دارای عناصری به صورت زیر است :

$$u_i(b) = y_i - x'_i b \quad i = 1, 2, \dots, n \quad (7-2)$$

در این حالت معادله $(5-2)$ به صورت زیر درمی‌آید و صدکهای رگرسیونی برحسب b از مینیمم کردن عبارت زیر به دست می‌آید:

$$S_w(b) = \sum_{i=1}^n r_w(u_i(b)).$$

تابع $S_w(b)$ به طور اکید محدب برحسب b و دارای مشتق پیوسته در b است. زمانی که هر مولفه b به بینهایت میل کند، مقدار تابع هم به بینهایت میل می‌کند. بنابراین $\beta_{RP\omega}$ منحصراً از حل معادله زیر به دست می‌آید:

$$S'_w(b) = \frac{\partial S_w(b)}{\partial b} = \begin{pmatrix} \vdots \\ \frac{\partial S_w(b)}{\partial b_j} \\ \vdots \end{pmatrix} = 0 \quad (7-2)$$

برای بیان ساده معادلات فوق، تابع پله‌ای زیر را تعریف

$$W(\lambda) = \begin{cases} 1 & \lambda \leq 0 \\ \omega & \lambda > 0 \end{cases} \quad (8-2)$$

با تعریف این تابع پله‌ای معادله $(4-2)$ تبدیل به رابطه:

$$r_\omega(\lambda) = W(\lambda)\lambda^2$$

می‌شود و با مشتق‌گیری از این رابطه داریم:

$$\frac{\partial r_\omega(\lambda)}{\partial \lambda} = 2W(\lambda)\lambda$$

که یک تابع پیوسته از λ می‌باشد. پس می‌توان نتیجه گرفت که بردار گرادیان $^{18} S'_w(b)$ یک تابع پیوسته از b است. بنابراین:

را برحسب β مینیمم می‌کنیم و حاصل را با $\beta_{RP\omega}$ نشان می‌دهیم.

یادآوری ۲-۱. در معادله $(5-2)$ اگر مقدار $\omega = 1$ باشد، آنگاه برآوردهای حداقل مربعات معمولی به دست می‌آید، یعنی:

$$\omega = 1 \implies \beta_{RP1} = \beta_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

یادآوری ۲-۲. به ازای تمام مقادیر ω ، آنچه صفحه $\hat{\mathcal{L}}^\omega$ بعدی را به وسیله $\beta_{RP\omega}$ در فضای P + بعدی (x_i, y_i) به صورت زیر تعریف می‌کنیم:

$$\hat{\mathcal{L}}_\omega \equiv \{Y = X\beta_{RP\omega}; X \in \mathbb{R}^P\}$$

که این را «صفحه ω – رگرسیونی»¹⁷ می‌نامیم. این صفحه به طور یکنواخت از پایین تا بالای انبوه داده‌ها با افزایش ω از صفر تا بینهایت حرکت می‌کند.

صفحه $\hat{\mathcal{L}}^\omega$ قرار می‌گیرند (یا به عبارت دیگر، نسبت تعداد نقاطی که ≤ 0 $\leq \omega$ ≤ 1 $\leq \alpha$ $\leq \omega$ $\leq P$ (به تعداد کل داده‌ها) تعريف می‌کنیم. برای $\alpha = P(\omega)$ (که مقادیری بین صفر و یک رامی گیرد) داریم :

$$\beta_{RP}^{(\alpha)} = \beta_{RP\omega}.$$

به عبارت دیگر برای مقدار مفروضی از α ، $\beta_{RP\omega}^{(\alpha)}$ برای مقادیری از ω است که بیانگر α درصد از داده‌هایی است که زیر صفحه $\hat{\mathcal{L}}^\omega$ قرار دارند. بنابراین سطح متناظر با $\beta_{RP}^{(\alpha)}$ را به صورت:

$$\hat{\mathcal{L}}^{(\alpha)} = \{Y = X\beta_{RP}^{(\alpha)}; X \in \mathbb{R}^P\} \quad (p(\omega) = \alpha)$$

تعريف می‌کنیم که در واقع 100α امین صدک رگرسیونی است. به عبارت دیگر صفحات مختلف $\hat{\mathcal{L}}^\omega$ را می‌توان به صورت

¹⁶ Hyperplane

¹⁷ ω -regression plane

¹⁸ Gradient vector

از طرفی مشتق مرتبه دوم (۹-۲) عبارت است از:

$$S''_{\omega}(\mathbf{b}) = \left(\frac{\partial^2 S_{\omega}(\mathbf{b})}{\partial b_j \partial b_n} \right)_{j,n=1,2,\dots,p} = 2 \mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{X} \quad (13-2)$$

اگر هیچیک از مقادیر $(\mathbf{b})_i$ صفر نباشد، آنگاه مشتق دوم حتماً پیوسته است.

حال اگر برای حل معادله (۷-۲) با یک مقدار آزمایشی \mathbf{b} ، و با توجه به معادله (۱۳-۲) از روش نیوتون - رافسون استفاده نماییم، یک مقدار \mathbf{b}_{NEW} به دست می آید، که می توان گفت با معادله:

$$\mathbf{b}_{NEW} = \beta(\omega, \mathbf{b}) \quad (14-2)$$

هم ارز است. بنابراین با روش تکراری (۱۴-۲) برای حل رابطه (۱۲-۲) مقادیر $\mathbf{b}_1 = \beta(\omega, \mathbf{b}_1)$ ، $\mathbf{b}_2 = \beta(\omega, \mathbf{b}_1)$ ، ... همان مقادیری هستند که رابطه $(\mathbf{b})_i$ را با روش نیوتون - رافسون مینیمم می کنند.

روش دیگر که برای به دست آوردن پاسخ $\beta_{RP\omega}$ در (۱۲-۲) می توان پیشنهاد کرد، بر اساس تغییرات ω به جای تغییرات \mathbf{b} است که به صورت زیر قابل استفاده می باشد. معادله (۱۱-۲) را بصورت زیر در نظر می گیریم:

$$\begin{aligned} \delta(\omega, \mathbf{b}) &= \beta(\omega, \mathbf{b}) - \mathbf{b} \\ &= [\mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{U}(\mathbf{b}) \end{aligned}$$

از طرفی:

$$\frac{\partial W(u_j(\mathbf{b}))}{\partial \omega} = I_+(\mathbf{u}_i(\mathbf{b}))$$

که:

$$I_+(\lambda) = \begin{cases} 0 & \text{if } \lambda \leq 0 \\ 1 & \text{if } \lambda > 0 \end{cases}$$

$$S'_{\omega}(\mathbf{b}) = -2 \mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{U}(\mathbf{b}) \quad (9-2)$$

که:

$$\mathbf{W}(\mathbf{U}(\mathbf{b})) = \text{diag}[W(u_i(\mathbf{b}))]$$

یک ماتریس قطری $n \times n$ است که عنصر قطر اصلی را $(\mathbf{W}(u_i(\mathbf{b})))$ تشکیل می دهد. از ترکیب روابط (۹-۲) و (۷-۲) نتیجه گرفته می شود که $\beta_{RP\omega}$ از حل معادله زیر به دست می آید:

$$\mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{U}(\mathbf{b}) = 0 \quad (10-2)$$

که برای حل معادله باد شده احتیاج به روش های تکراری ^{۱۹} داریم. از این معادله می توان نتیجه گرفت:

$$\mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{Y} = \mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{X} \mathbf{b}$$

و اگر پاسخ رابطه فوق را با (\mathbf{b}, ω) نشان دهیم، خواهیم داشت: [۱۰]

$$\beta(\omega, \mathbf{b}) = [\mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{Y}$$

قابل توجه است که وقتی وزنهای $(\mathbf{u}_i(\mathbf{b}))$ ثابت باشند، پاسخ گاوس مارکف ^{۲۰} برای انتخاب β در جهت مینیمم کردن مجموع مربعات :

$$\sum_{i=1}^n W(u_i(\mathbf{b}))(y_i - \mathbf{x}'_i \beta)^2$$

آنگاه $\beta(\omega, \mathbf{b})$ است:

$$\beta(\omega, \mathbf{b}) - \mathbf{b} = [\mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}(\mathbf{U}(\mathbf{b})) \mathbf{U}(\mathbf{b}) \quad (11-2)$$

و این مقدار با توجه به معادله (۱۰-۲) برابر با صفر است. بنابراین با توجه به خاصیت مانایی ^{۲۱} داریم:

$$\beta(\omega, \beta_{RP\omega}) = \beta_{RP\omega} \quad (12-2)$$

مستقل وجود ندارد، (یعنی برای هر $x_i = 1$ ، وقتی $n \rightarrow \infty$) بافتنگار y_1, y_2, \dots, y_n به سمت توزیع $(\alpha, 1)$ میل می‌کند. در این وضعیت $\beta_{RP\omega}$ را با β_ω نشان می‌دهیم. ω بیانگر مقدار ω است که β_ω برابر با $Z^{(\alpha)}$ باشد. $(\Phi^{-1}(\alpha))^{(\alpha)} = Z^{(\alpha)}$ نقطه 100α امین درصد از توزیع نرمال استاندارد است.

$$\omega^{(\alpha)} = 1 + \frac{Z^{(\alpha)}}{[\phi(Z^{(\alpha)}) - (1 - \alpha)Z^{(\alpha)}]} \quad (18-2)$$

که در این رابطه $\phi(Z) = e^{-\frac{1}{2}z^2}/\sqrt{2\pi}$ و داریم $(\frac{1}{\omega^{(\alpha)}})^{(\alpha)} = \omega^{(\alpha)}$. که با توجه به این فرمول مقادیر زیر را به دست می‌آوریم:

$$\begin{aligned} \alpha : & \quad 0/5 \quad 0/67 \quad 0/75 \quad 0/84 \quad 0/90 \quad 0/95 \\ \omega^{(\alpha)} : & \quad 1 \quad 2/02 \quad 5/02 \quad 12/81 \quad 28/11 \quad 80/04 \quad 19-2 \end{aligned}$$

این مقادیر را می‌توان با مقادیری که از محاسبات کلسترومین به دست آمده مقایسه کرد. مثلاً $\hat{\omega}(0/75) = 6/31 \approx 2/9$ ، $\hat{\omega}(0/67) = 24/8 \approx 3/4$ ، $\hat{\omega}(0/84) = 14/3$.

با تغییر دادنتابع چگالی نرمال استاندارد $\phi(z)$ به صورت:

$$\phi^{(0/75)}(z) = \begin{cases} \phi(z)/c & z \leq z^{(0/75)} = 0/674 \\ 5/02\phi(z)/c & z > 0/674 \end{cases} \quad (20-2)$$

مقدار ثابت c را طوری انتخاب می‌کنیم که $\phi^{(0/75)}(z)$ یک تابع چگالی احتمال تبدیل شود. آنگاه $\phi^{(0/75)}(z)$ دارای میانگین زیراست:

$$\int_{-\infty}^{+\infty} z\phi^{(0/75)}(z)dz = 0/675 \quad (21-2)$$

که این مطلب در خصوص ناوردایی با معادله (۱۴-۲) در ارتباط است. نکته جالب توجه این است که $\phi^{(\alpha)}$ دارای واریانس ۱ می‌باشد یعنی:

$$\phi^{(\alpha)}(z) \sim (z^{(\alpha)}, 1)$$

تعریف می‌کنیم. حال (b, ω) را تابعی از ω با مقدار ثابت b در نظر می‌گیریم. آنگاه خواهیم داشت:

$$\frac{\partial \delta(\omega, b)}{\partial \omega} = [\mathbf{X}' \mathbf{W}(\mathbf{U}(b)) \mathbf{X}]^{-1} \left\{ \mathbf{X}' \mathbf{I}_+(\mathbf{U}(b)) \mathbf{U}(b) - \frac{\partial [\mathbf{X}' \mathbf{W}(\mathbf{U}(b)) \mathbf{X}]^{-1}}{\partial \omega} \delta(\omega, b) \right\}$$

که $(\mathbf{U}(b)) \mathbf{I}_+(\mathbf{U}(b))$ یک ماتریس قطری $n \times n$ با درایه قطری $u_i(b)$ می‌باشد. با توجه به رابطه فوق می‌توان نشان داد:

$$\frac{\partial \beta_{RP\omega}}{\partial \omega} = \frac{1}{1 + \omega} [\mathbf{X}' \mathbf{W}(\mathbf{U}(\beta_{RP\omega})) \mathbf{X}]^{-1} \mathbf{X}' |\mathbf{U}(\beta_{RP\omega})| \quad (15-2)$$

که $|\mathbf{U}(b)|$ برداری از قدر مطلق باقیمانده‌ها به صورت $|u_1(b)|, \dots, |u_n(b)|$ است. با پیدا کردن $\beta_{RP\omega}$ برای بعضی از مقادیر ω می‌توان پاسخ را در یک همسایگی ω به فرم $\omega + \Delta\omega$ به صورت زیر تقریب زد:

$$\beta_{RP(\omega+\Delta\omega)} = \beta_{RP\omega} + \frac{\partial \beta_{RP\omega}}{\partial \omega} \cdot \Delta\omega \quad (16-2)$$

صدکهای رگرسیونی توسط یک الگوریتم که قدمهای (۱۶-۲) و (۱۴-۲) را تکرار می‌کند پیدا می‌شود. با شروع از یک مقدار دو مقدار جدید زیر به دست می‌آید و این عمل تکرار می‌شود.

$$\begin{aligned} \beta_{RP(\omega+\Delta\omega)}^{(0)} &= \beta_{RP} + \frac{\partial \beta_{RP\omega}}{\partial \omega} \cdot \Delta\omega \\ &\quad \& \\ \beta_{RP(\omega+\Delta\omega)}^{(1)} &= \beta(\omega + \Delta\omega, \beta_{(\omega+\Delta\omega)}^{(0)}) \end{aligned} \quad (17-2)$$

هرچه $\Delta\omega$ کوچکتر باشد، $\beta_{RP(\omega+\Delta\omega)}^{(1)}$ تقریب خوبی برای $\beta_{RP(\omega+\Delta\omega)}$ می‌باشد. ω برای قدم بعدی عبارت است از $\omega + \Delta\omega$ و برای شروع محاسبات بهتر است $\omega = 1$ در نظر گرفته شود. در این صورت داریم:

$$\beta_{RP1} = [\mathbf{X}' \mathbf{X}]^{-1} \mathbf{X}' \mathbf{Y}$$

که در واقع برآورد مینیمم مربعات خطاهاست. یادآوری ۳-۲. یک ایده کلی در مورد رابطه بین ω و α ، حالت خاص زیر است. در این حالت فرض می‌کنیم متغیر

۳ یک مثال تجربی

اطلاعات مربوط به قندخون ناشتایی (FBS)^{۲۵} و هموگلوبین ($A1$)^{۲۶} را به عنوان یک پارامتر در خون که نشان دهنده وضعیت بیماری دیابت^{۲۷} است، در نظرمی گیریم.^{۲۸} x (قندخون ناشتایی) به عنوان متغیر مستقل و y (هموگلوبین $A1$) به عنوان متغیر وابسته برای $n = 216$ نفر از بیمارانی است که زیر نظر قرار گرفته‌اند.

در شکل (۱-۲) نمودار پراکندگی (x_i, y_i) را همراه با خطوط صد کهای رگرسیونی که براساس روش (۱-۱) به دست آمده، مشاهده می‌کنیم که برای مقادیر $\alpha = ۰/۹۵, ۰/۱۰, ۰/۳۳, ۰/۵۰, ۰/۶۷, ۰/۹۰, ۰/۹۵$ رسم شده است.

در شکل (۲-۳) نمودار پراکندگی (x_i, y_i) همراه با خطوط صد کهای رگرسیونی برای $\omega = ۰/۰۵, \omega = ۰/۰۵, \omega = ۱, \omega = ۲, \omega = ۲۵$ رسم شده است.

با توجه به شکل (۱-۳) مشاهده می‌شود اگر آزمایش قندخون ناشتایی فردی، با تمرکز ۹۰٪ برابر با ۱۰۰ باشد انتظار می‌رود هموگلوبین $A1$ آن حدود ۶/۲ تا ۸/۷ باشد.

درجول (۱-۳) برآورد صد کهای رگرسیونی را برای ω های مختلف به دست آورده و رابطه آن را با $\alpha = p(\omega)$ نشان داده‌ایم. در این جدول برای مقادیر مختلف ω ضرایب رگرسیونی و α های متناظر آن محاسبه شده و همانطور که مشاهده می‌شود ضریب قندخون ناشتایی یک روند مثبت دارد بدین معنی که با افزایش آن مقدار هموگلوبین $A1$ افزایش پیدا می‌کند.

برای تمام مقادیر α ، $1 < \alpha < \phi$ این رابطه برقرار است. بنابراین ϕ شبیه به خانواده مکانی^{۲۹} عمل می‌کند. این نتیجه مخصوص توزیع نرمال است.

یادآوری ۴-۲. تابع زیان مربع خطای نامتقارن (۴-۲) خاصیت مهم ناوردایی را نتیجه می‌دهد و اگر بردار مشاهدات (y_1, \dots, y_n) به (cy_1, \dots, cy_n) تغییر کند آنگاه پاسخ رابطه (۵-۲) از $c\beta_{RP\omega}$ به $c\beta_{RP\omega} + (d, 0, 0, \dots, 0)$ تغییر خواهد کرد. بنابر قضیه ۱ که توسط نیوی و پاول [۸] گفته شد، برآورد $\beta_{RP\omega}$ دارای خاصیت ناوردایی در مکان و مقیاس^{۲۹} است و داریم:

$$y \rightarrow cy + d \mathbf{1} \implies \beta_{RP\omega} \rightarrow c\beta_{RP\omega} + (d, 0, 0, \dots, 0)'$$

بدون داشتن خاصیت ناوردایی در مقیاس، برآورد صد کهای رگرسیونی به مقیاسی که از آن استفاده شده بستگی خواهد داشت.

ناوردایی مقیاس فقط منحصر به تابع زیان مربع خطای نامتقارن نیست، بلکه هرتابع زیان توانی^{۲۹} به صورت:

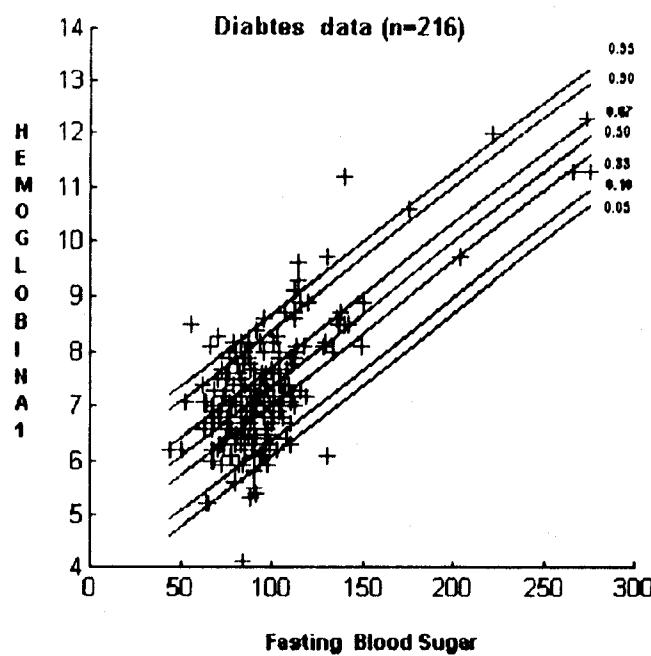
$$r_\omega(\lambda) = \begin{cases} |\lambda|^p & \lambda \leq 0 \\ \omega|\lambda|^p & \lambda > 0 \end{cases}$$

ناوردایی مقیاس را برای برآوردگر متناظر (یعنی β_ω) نتیجه می‌دهد. از رابطه فوق به سادگی $r_\omega(c\lambda) = |c|^p r_\omega(\lambda)$ نتیجه می‌شود که در میان تمام حالت‌های $p > ۰$ ، حالت (۲) دارای مزیت محاسباتی می‌باشد.

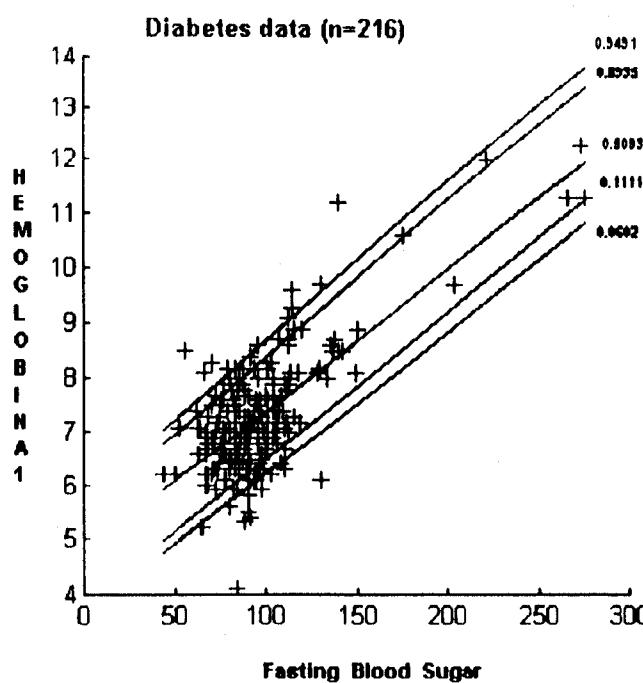
یادآوری ۵-۲. خاصیت سازگاری برآوردگرهای صد کهای رگرسیونی توسط نیوی و پاول (۱۹۸۷)[۸] بیان شده است. آنها در لم-A نتایج کلی سازگاری در همگرایی صد کهای رگرسیونی را به متناظر شان در جمعیت بیان می‌کنند.

Location Family—	^{۲۹}
Scale and location invariance —	^{۲۹}
Power loss function—	^{۲۹}
Fasting Blood Sugar—	^{۲۵}
Hemoglobin A1—	^{۲۶}
Diabetes—	^{۲۷}

^{۲۸} این اطلاعات توسط آقای دکتر امینی رئیس مرکز تحقیقات غدد و متابولیسم بیمارستان امین اصفهان جمع آوری شده است.



شکل ۱-۳ . این شکل مربوط به بیماران دیابتی است که اطلاعات مربوط به قندخون ناشتابی (FBS) و هموگلوبین $A1$ ($HbA1$) آنها برای زوجهای مرتب (x_i, y_i) در مقابل هم، و نیز منحنی حداقل مربعات معمولی (OLS) برآورده شده به صورت $y_i = b_0 + b_1 x_i + e_i$ همراه با صدکهایی که از روش ۱-۱ برای مقادیر $\alpha = ۰.۵۰, ۰/۰۵, ۰/۱۰, ۰/۲۲, ۰/۵۰, ۰/۶۷, ۰/۹۰, ۰/۹۵$ رسم شده است.



شکل ۲-۳ . این شکل مربوط به بیماران دیابتی است که اطلاعات مربوط به قندخون ناشتابی (FBS) و هموگلوبین $A1$ ($HbA1$) آنها برای زوجهای مرتب (x_i, y_i) در مقابل هم، و نیز صدکهایی رگرسیونی برای $\omega = ۱, \omega = ۰/۰۵, \omega = ۰/۰۲, \omega = ۰/۰۵, \omega = ۰/۷۲, \omega = ۰/۲۵$ رسم شده است.

جدول ۳-۱- تحلیل صدکهای رگرسیونی برای بیماران دیابتی

w	(β_1 , β_2)	$p(w)$
۰/۰۲	(۲/۵۸۴۰ , ۰/۰۲۶۳)	۰/۰۷۰۲
۰/۰۵	(۲/۷۷۸۴ , ۰/۰۲۷۱)	۰/۱۱۱۱
۰/۰۸	(۲/۹۰۲۶ , ۰/۰۲۶۸)	۰/۱۴۸۱
۰/۱	(۴/۰۲۴۷ , ۰/۰۲۶۶)	۰/۱۷۵۹
۰/۱۱	(۴/۰۶۹۷ , ۰/۰۲۶۶)	۰/۱۸۰۷
۰/۱۲	(۴/۱۰۱۹ , ۰/۰۲۶۵)	۰/۱۸۹۸
۰/۱۵	(۴/۱۸۲۲ , ۰/۰۲۶۴)	۰/۲۱۲۰
۰/۲	(۴/۴۱۸۲ , ۰/۰۲۶۰)	۰/۳۲۸۷
۰/۵	(۴/۵۷۲۴ , ۰/۰۲۶۰)	۰/۴۰۲۸
۱	(۴/۷۷۱۲ , ۰/۰۲۶۱)	۰/۵۰۹۳
۱/۵	(۴/۸۸۱۹ , ۰/۰۲۶۲)	۰/۵۶۹۴
۲	(۴/۹۰۷۴ , ۰/۰۲۶۴)	۰/۶۲۴۲
۲/۲	(۴/۹۸۲۴ , ۰/۰۲۶۴)	۰/۶۴۲۵
۹	(۰/۲۸۲۶ , ۰/۰۲۸۱)	۰/۷۹۱۷
۱۰	(۰/۳۰۲۱ , ۰/۰۲۸۲)	۰/۸۱۰۲
۱۲	(۰/۳۴۰۲ , ۰/۰۲۸۳)	۰/۸۲۴۱
۱۳	(۰/۳۵۷۱ , ۰/۰۲۸۴)	۰/۸۲۴۱
۲۰	(۰/۵۲۶۲ , ۰/۰۲۸۶)	۰/۸۹۳۵
۲۷	(۰/۵۴۹۳ , ۰/۰۲۸۷)	۰/۸۹۳۵
۲۹	(۰/۵۶۹۹ , ۰/۰۲۸۷)	۰/۸۹۸۱
۳۰	(۰/۵۷۸۷ , ۰/۰۲۸۶)	۰/۹۰۲۸
۵۰	(۰/۷۰۴۰ , ۰/۰۲۸۹)	۰/۹۴۴۴
۵۵	(۰/۷۲۷۹ , ۰/۰۲۹۰)	۰/۹۴۴۴
۷۰	(۰/۷۵۱۱ , ۰/۰۲۹۰)	۰/۹۴۹۱
۷۵	(۰/۷۷۲۰ , ۰/۰۲۹۱)	۰/۹۴۹۱

مراجع

- [1] D.Aigner, T.Amemiya, and D.Poirier,"On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function." *Internat. Econom. Rev.* 17,(1976) 372-396.
- [2] J.Breckling, and R.Chambers, "M-quantiles." *Biometrika* 75, (1988) 761-771.
- [3] B. Efron, "Regression Percentiles Using Asymmetric Squared Error Loss." *Statistica Sinica* 1.(1991) 93-125.
- [4] H. Glejser, "A new test for heteroskedasticity." *J. Amer. Statist. Assoc.* 64,(1969) 316-323.
- [5] P.J. Huber, "Robust Statistics." *John Wiley, New York.*(1981)

- [6] R. Koenker, and G. Bassett, "Regression quantiles." *Econometrica* 46,(1978) 33-50.
- [7] Lipid Research Clinics Program "The Lipid Research Clinics Coronary Primary Prevention Trial Results (Pars I and II)." *J. Amer. Med. Assoc.* 251,(1984) 351-374.
- [8] W. K. Newey, and J.L. Powell, "Asymmetric least squares estimation and testing." *Econometrica* 55,(1987) 819-847.
- [۹] جمشیدیان، احمد رضا: «صد کهای رگرسیونی با استفاده ازتابع زیان مربع نامتقارن خطأ» پایان نامه (۱۳۷۷) دانشگاه صنعتی اصفهان
- [۱۰] کتوور: «آمارنپارامتری کاربردی » چاپ اول (۱۳۷۲) مرکز نشر دانشگاهی تهران. مترجم: هاشمی پرست، سید مقتدی
-