صدکهای رگرسیونی با استفاده‌از تابع زیان مربع خطای نامتقارن

احمدرضا جمشیدیان   سروش علیمرادی[۱]

چکیده

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

۱ مقدمه

داده‌هائی که برای انجام یک تحلیل رگرسیونی به کارمی‌روند به فرم انبوهی ازنقاط در فضای اقلیدسی به صورت زیر هستند:

$$(x_i, y_i) \qquad i = ۱, ۲, ۳, \cdots, n$$

که $\mathbf{x}_i$ یک بردار $p \times ۱$ از متغیرهای مستقل و y_i یک مقدارحقیقی از متغیر پاسخ است. خروجی اولیه یک تحلیل معمولی برآورد تابع رگرسیونی است که نقاط وسط انبوه، نقاط را نمایش می دهند. اگر علاقه‌مند به شرح نقاط پایینی یا بالایی به همان خوبی نقاط میانی باشیم، چگونه عمل می کنیم؟

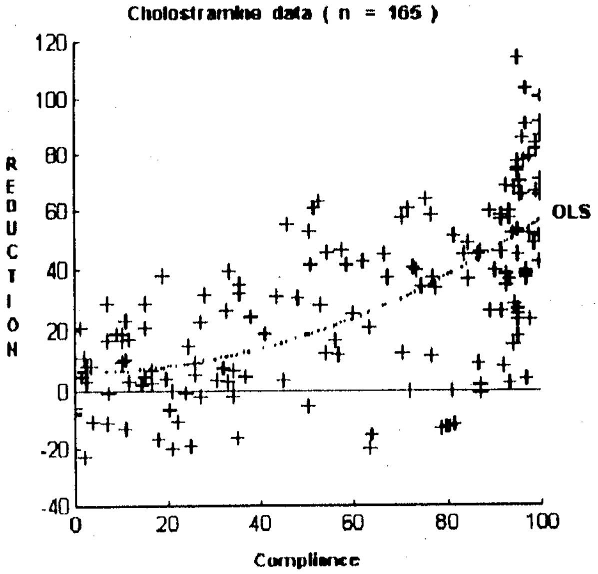
افرون[۲] (۱۹۹۱)[۳] دراین خصوص تحقیقاتی را انجام داده که مقاله حاضر برگرفته از مقاله اوست. در اینجا سعی بر این است که روش مفیدی را برای برآورد صدکهای رگرسیونی شرح دهیم . برای مثال ۲۵ اُمین یا ۷۵ اُمین صدک شرطی Y به عنوان تابعی از X برآورد می شوند.

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

۱— احمدرضا جمشیدیان، سروش علیمرادی، گروه آمار دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان

۲— B. Efron

داروی کلسترومین[۳] و تأثیر آن بر پایین آوردن کلسترول خون[۴] است. بنابر این y_i میزان کاهش[۵] خون و x_i[۶] درصد بسته‌های کلسترومین مصرفی است که بیمار i اُم قبول کرده است. منحنی حداقل مربعات معمولی برازانده 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 کلسترومین و تاثیر آن بر پایین آوردن کلسترول خون، همراه با منحنی حداقل مربعات معمولی برازانده شده بصورت مدل رگرسیونی درجه دو، $y_i = b_0 + b_1 x_i + b_2 x_i^2 + e_i$ است.

شکل (۱—۲) صدکهای رگرسیونی این داده‌ها را نشان می‌دهد که به وسیلهٔ روش حداقل مربعات نامتقارن، که در اینجا مورد نظر است، برآورد شده‌اند. صدکهای رگرسیونی اطلاعات بیشتری نسبت به خط حداقل مربعات معمولی در اختیار ما قرار می‌دهند. در این آزمایش اگر بیماری ۶۰ درصد از دارویی را که قبول کرده، مصرف کند، با تمرکز ۸۰ درصد پاسخ، ۱۰ اُمین تا ۹۰ اُمین صدک برآورد می‌شود تا یک کاهش کلسترول خون مابین ۶— تا ۵۴ 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 $e_i \sim N(0, \sigma^2)$ برای $i = 1, 2, 3, \cdots, n$) باشد، برآورد σ^2 را بصورت $\hat{\sigma}^2$ نشان می دهیم، آنگاه برآورد α ۱۰۰ اُمین صدک رگرسیونی به صورت :

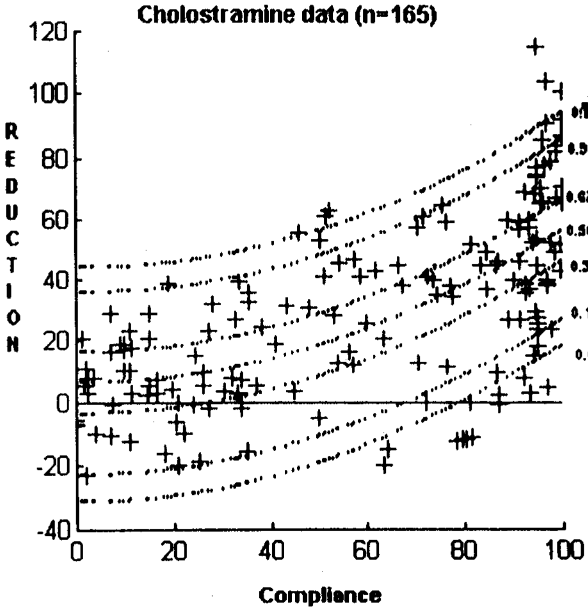
$$\hat{\mu}(x) + \hat{\sigma} Z^{(\alpha)} \qquad (1-1)$$

به دست می آید[۹] که $\hat{\mu}(x)$ برآورد رگرسیون حداقل مربعات معمولی و $Z^{(\alpha)}$ هم α ۱۰۰ اُمین صدک از تابع چگالی نرمال استاندارد $Z^{(\alpha)} = \Phi^{-1}(\alpha)$ است. اگر تمام این خطوط را برای α های متفاوت دریک صفحه محورهای مختصات رسم کنیم تمام آنها موازی و فاصله آنها $\hat{\sigma} Z^{(\alpha)}$ است.

ایدهٔ اصلی صدکهای رگرسیونی بر اساس ابداعات کانکر و باست[۸] (۱۹۷۸) [۶] می‌باشد. آنها از قدرمطلق زیان نامتقارن[۹] بیشتر از مربع زیان نامتقارن[۱۰] استفاده کرده، در این بین چندکهای رگرسیونی[۱۱] را تعریف کُردند. پاول و نیوی[۱۲] (۱۹۸۷) [۸] کاری را که توسط آیگنر، آممیا و پویریر[۱۳] (۱۹۷۶) [۱] انجام شده بود، ادامه دادند و از مربع

۳— Cholostyramine
۴— Blood Cholesterol
۵— Reduction
۶— Compliance
۷— این اطلاعات مربوط به بیماران تحت کنترل بخش استانفورد می‌باشد که در *LipidRsearchClinicsProgram* (۱۹۸۴) [۷] توسط *Drs.D.FeldmanandJ.Farquhar* از دانشکدهٔ داروسازی استانفورد جمع آوری شده است.
۸— R. Koenker and G. Bassett
۹— Asymmetric Absolute Loss
۱۰— Asymmetric Squared Loss
۱۱— Regression Quantiles
۱۲— Powell and Newey
۱۳— Aigner , Amemiya and poirier

خطای نامتقارن آنچنان که ما در اینجا از آن استفاده خواهیم کرد، استفاده کردند و نتایج منحنی را انتظارک [۱۴] 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.



شکل ۱–۲: این شکل مربوط به صدکهای رگرسیونی‌متداول برای آزمایش کلینیکی در استفاده از داروی کلسترومین و تاثیر آن بر پایین آوردن کلسترول خون است و برای $\alpha = ۰.۰۵, ۰.۱۰, ۰.۳۳, ۰.۵۰, ۰.۶۷, ۰.۹۰, ۰.۹۵$ صدکهای رگرسیونی آن رسم شده است.

دراینجا صدکهای رگرسیونی که اولین بار توسط افرون (۱۹۹۱)[۳] ارایه شد جالبترند. وی صدکهای نمونه‌ای را که در توصیف اطلاعات نمونه‌ای مفیدند، براساس چندکهای رگرسیونی تعریف شده توسط کانکر و باست[۶] به رگرسیون خطی تعمیم داد و آن را صدکهای رگرسیونی نامید. این صدکها مفسرهای 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.

۲ صدکهای رگرسیونی

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

$$y_i = \beta_۱ x_{i۱} + \beta_۲ x_{i۲} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad (۱-۲)$$

که فرم برداری آن به صورت زیر است :

$$y_i = x_i'\beta + \varepsilon_i \quad x_{i۱} = ۱ \quad \& \quad \varepsilon_i \sim (۰, \sigma^۲) \quad i = ۱, ۲, \ldots, n \qquad (۲-۲)$$

و فرم ماتریس مدل هم به صورت زیر می‌باشد:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (۳-۲)$$

که $\varepsilon = (\varepsilon_۱, \ldots, \varepsilon_n)'$، بردار خطایی است دارای تابع توزیع F با میانگین صفر و واریانس ثابت $\sigma^۲$ ، $\beta = (\beta_۱, \ldots, \beta_n)'$ بردار پارامترهای مجهول، $\mathbf{X}$ یک ماتریس $n \times p$ معلوم بارتبه کامل ستونی و $\mathbf{Y} = (y_۱, \ldots, y_n)'$ بردار مشاهدات است.

برای برآورد صدکهای رگرسیونی پارامتر β، ابتدا تابع زیان مربع خطای نامتقارن را برای هر $\omega > ۰$ به فرم:

$$r_\omega(\lambda) = \begin{cases} \lambda^۲ & \lambda \leq ۰ \\ \omega\lambda^۲ & \lambda > ۰ \end{cases} \qquad (۴-۲)$$

تعریف کرده و سپس معادله:

$$S_\omega(\beta) = \sum_{i=۱}^{n} r_\omega(y_i - x_i'\beta) \qquad (۵-۲)$$

[۱۴] Expectile
[۱۵] Robustify

را برحسب β مینیمم می‌کنیم و حاصل را با β_{RP_ω} نشان می‌دهیم.

یادآوری ۲–۱. در معادله (۲–۵) اگر مقدار $\omega = 1$ باشد، آنگاه برآوردگر حداقل مربعات معمولی به دست می‌آید، یعنی:

$$\omega = 1 \Longrightarrow \boldsymbol{\beta}_{RP1} = \boldsymbol{\beta}_{OLS} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

یادآوری ۲–۲. به ازای تمام مقادیر $\omega > 0$، اَبَرصفحه[۱۶] p بعدی $\hat{\mathcal{L}}_\omega$ را به وسیلهٔ $\boldsymbol{\beta}_{RP_\omega}$ در فضای $P+1$ بعدی $(\mathbf{x}_i, y)$ به صورت زیر تعریف می‌کنیم:

$$\hat{\mathcal{L}}_\omega \equiv \{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_{RP_\omega}; \boldsymbol{X} \in \Re^P\}$$

که این را «**صفحه ω – رگرسیونی**»[۱۷] می‌نامیم. این صفحه به طور یکنواخت از پایین تا بالای انبوه داده‌ها با افزایش ω از صفر تا بینهایت حرکت می‌کند.

$P(\omega)$ را نسبت تعداد نقاط (x_i, y_i) به کل داده‌ها که زیر صفحه $\hat{\mathcal{L}}_\omega$ قرار می‌گیرند (یا به عبارت دیگر، نسبت تعداد نقاطی که $u_i(\boldsymbol{\beta}_{RP_\omega}) \leq 0$، (۲–۶) به تعداد کل داده‌ها) تعریف می‌کنیم. برای $P(\omega) = \alpha$ (که α مقادیری بین صفر و یک را می‌گیرد) داریم :

$$\boldsymbol{\beta}_{RP}^{(\alpha)} = \boldsymbol{\beta}_{RP_\omega} .$$

به عبارت دیگر برای مقدارمفروضی از α، $0 < \alpha < 1$، بردار $\boldsymbol{\beta}_{RP}^{(\alpha)}$ همان بردار $\boldsymbol{\beta}_{RP_\omega}$ برای مقادیری از ω است که بیانگر α درصد از داده‌هایی است که زیر صفحه $\hat{\mathcal{L}}_\omega$ قرار دارند. بنابراین سطوح متناظر با $\boldsymbol{\beta}_{RP}^{(\alpha)}$ را به صورت:

$$\hat{\mathcal{L}}^{(\alpha)} = \{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_{RP}^{(\alpha)}\ ;\ \boldsymbol{X} \in \Re^P\} \qquad (p(\omega) = \alpha)$$

تعریف می کنیم که در واقع 100α اُمین صدک رگرسیونی است. به عبارت دیگرصفحات مختلف $\hat{\mathcal{L}}_\omega$ را می‌توان به صورت $\hat{\mathcal{L}}^{(\alpha)}$ نیز نشان داد واین نامگذاری براساس نسبت مشاهدات $(\mathbf{x}_i, y_i)$ که پایین این صفحات واقع شده‌اند، می‌باشد.

برای محاسبهٔ صدکهای رگرسیونی، اگر فرض کنیم $\mathbf{b}$ یکی از مقادیر مورد نظر برای معادلهٔ (۲–۲) باشد، آنگاه بردار باقی مانده‌ها $\mathbf{U(b)}$ دارای عناصری به صورت زیر است :

$$u_i(\boldsymbol{b}) = y_i - \boldsymbol{x}_i'\boldsymbol{b} \qquad i = 1, 2, \ldots, n \qquad (۲-۶)$$

در این حالت معادله(۲–۵) به صورت زیر درمی‌آید و صدکهای رگرسیونی برحسب $\mathbf{b}$ از مینیمم کردن عبارت زیر به دست می‌آید:

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

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

$$S_w'(\boldsymbol{b}) = \frac{\partial S_w(\boldsymbol{b})}{\partial \boldsymbol{b}} = \begin{pmatrix} \vdots \\ \frac{\partial S_w(\boldsymbol{b})}{\partial b_j} \\ \vdots \end{pmatrix} = 0 \qquad (۲-۷)$$

برای بیان ساده معادلات فوق، تابع پله‌ای زیر را تعریف می‌کنیم:

$$W(\lambda) = \begin{cases} 1 & \lambda \leq 0 \\ \omega & \lambda > 0 \end{cases} \qquad (۲-۸)$$

با تعریف این تابع پله‌ای معادله (۲–۴) 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[۱۸] $S_w'(\mathbf{b})$ یک تابع پیوسته از $\mathbf{b}$ است. بنابراین:

---
[۱۶] Hyperplane
[۱۷] ω-regression plane
[۱۸] Gradient vector

$$S'_\omega(b) = -2X'W(U(b))U(b) \qquad (۲-۹)$$

که:

$$W(U(b)) = diag[W(u_i(b))]$$

یک ماتریس قطری $n \times n$ است که عناصر قطر اصلی را $W(u_i(\mathrm{b}))$ ها تشکیل می‌دهد. از ترکیب روابط (۲−۹) و(۲−۷) نتیجه گرفته می‌شود که β_{RP_ω} از حل معادلهٔ زیر به دست می‌آید:

$$X'W(U(b))U(b) = 0 \qquad (۲-۱۰)$$

که برای حل معادلهٔ یاد شده احتیاج به روشهای تکراری[۱۹] داریم. از این معادله می‌توان نتیجه گرفت:

$$X'W(U(b))Y = X'W(U(b))Xb$$

و اگر پاسخ رابطهٔ فوق را با $\beta(\omega, b)$ نشان دهیم، خواهیم داشت:[۱۰]

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

قابل توجه است که وقتی وزنهای $W(u_i(\mathrm{b}))$ ثابت باشند، پاسخ گاوس مارکف[۲۰] برای انتخاب β در جهت مینیمم کردن مجموع مربعات :

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

$\beta(\omega, b)$ است، آنگاه :

$$\beta(\omega, b) - b = [X'W(U(b))X]^{-1}X'W(U(b))U(b) \qquad (۲-۱۱)$$

و این مقدار با توجه به معادلهٔ (۲−۱۰) برابر با صفر است. بنابراین با توجه به خاصیت مانایی[۲۱] داریم:

$$\beta(\omega, \beta_{RP_\omega}) = \beta_{RP_\omega} \qquad (۲-۱۲)$$

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

$$S''_\omega(b) = \left(\frac{\partial^2 S_\omega(b)}{\partial b_j \partial b_n}\right)_{j,n=1,2,\dots,p} = 2X'W(U(b))X \qquad (۲-۱۳)$$

اگر هیچیک ازمقادیر $u_i(\mathrm{b})$ صفرنباشد، آنگاه مشتق دوم حتماً پیوسته است.

حال اگر برای حل معادلهٔ (۲−۷) با یک مقدار آزمایشی b ، و با توجه به معادلهٔ (۲−۱۳)از روش نیوتن - رافسون استفاده نماییم، یک مقدار b_{NEW} به دست می‌آید، که می‌توان گفت با معادلهٔ:

$$b_{NEW} = \beta(\omega, b) \qquad (۲-۱۴)$$

هم ارز است. بنابراین با روش تکراری (۲−۱۴) برای حل رابطهٔ (۲−۱۲) مقادیر $b_1 = \beta(\omega, b_0)$ ، $b_2 = \beta(\omega, b_1)$ ، ... همان مقادیری هستند که رابطهٔ $S_\omega(b)$ را با روش نیوتن- رافسون مینیمم می‌کنند.

روش دیگر که برای به دست آوردن پاسخ β_{RP_ω} در (۲−۱۲) می‌توان پیشنهاد کرد، بر اساس تغییرات ω به جای تغییرات b است که به صورت زیر قابل استفاده می باشد. معادلهٔ (۲−۱۱) را بصورت زیر در نظر می‌گیریم:

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

از طرفی:

$$\frac{\partial W(u_j(b))}{\partial \omega} = I_+(u_i(b))$$

که:

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

۱۹− Iterative methods
۲۰− Gauss Markov
۲۱− Stationary

تعریف می‌کنیم. حال $\delta(\omega, \mathbf{b})$ را تابعی از ω با مقدار ثابت $\mathbf{b}$ در نظر می‌گیریم. آنگاه خواهیم داشت:

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

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

$$\frac{\partial \boldsymbol{\beta}_{RP\omega}}{\partial \omega} = \frac{1}{1+\omega}[\boldsymbol{X'W}(\boldsymbol{U}(\boldsymbol{\beta}_{RP\omega}))\boldsymbol{X}]^{-1}\boldsymbol{X'}|\boldsymbol{U}(\boldsymbol{\beta}_{RP\omega})| \qquad (۲-۱۵)$$

که $|\mathbf{U}(\mathbf{b})|$ برداری از قدر مطلق باقی‌مانده‌ها به صورت $(|u_1(b)|, \cdots, |u_n(b)|)'$ است.

با پیدا کردن $\boldsymbol{\beta}_{RP\omega}$ برای بعضی از مقادیر ω می‌توان پاسخ را در یک همسایگی ω به فرم $\omega + \Delta\omega$ به صورت زیر تقریب زد:

$$\boldsymbol{\beta}_{RP(\omega+\Delta\omega)} = \boldsymbol{\beta}_{RP\omega} + \frac{\partial \boldsymbol{\beta}_{RP\omega}}{\partial \omega}.\Delta\omega \qquad (۲-۱۶)$$

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

$$\boldsymbol{\beta}^{(0)}_{RP(\omega+\Delta\omega)} = \boldsymbol{\beta}_{RP} + \frac{\partial \boldsymbol{\beta}_{RP\omega}}{\partial \omega}.\Delta\omega$$
$$\&$$
$$\boldsymbol{\beta}^{(1)}_{RP(\omega+\Delta\omega)} = \boldsymbol{\beta}(\omega+\Delta\omega, \boldsymbol{\beta}^{(0)}_{(\omega+\Delta\omega)}) \qquad (۲-۱۷)$$

هرچه $\Delta\omega$ کوچکتر باشد، $\boldsymbol{\beta}^{(1)}_{RP(\omega+\Delta\omega)}$ تقریب خوبی برای $\boldsymbol{\beta}_{RP(\omega+\Delta\omega)}$ 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:

$$\boldsymbol{\beta}_{RP1} = [\boldsymbol{X'X}]^{-1}\boldsymbol{X'Y}$$

که در واقع برآورد مینیمم مربعات خطاهاست.

یادآوری ۲-۳. یک ایده کلی در مورد رابطه بین ω و α، حالت خاص زیر است. دراین حالت فرض می کنیم متغیر مستقل وجود ندارد، (یعنی برای هر i، $x_i = 1$)، وقتی $n \to \infty$، بافت‌نگار $y_1, \cdots, y_n$ به سمت توزیع $N(0, 1)$ میل می‌کند. در این وضعیت $\boldsymbol{\beta}_{RP\omega}$ را با β_ω نشان می‌دهیم. $\omega^{(\alpha)}$ بیانگر مقدار ω است که β_ω برابر با $Z^{(\alpha)}$ باشد. ($Z^{(\alpha)} = \Phi^{-1}(\alpha)$ نقطه 100α اُمین درصد از توزیع نرمال استاندارد است.)

$$\omega^{(\alpha)} = 1 + \frac{Z^{(\alpha)}}{[\phi(Z^{(\alpha)}) - (1-\alpha)Z^{(\alpha)}]} \qquad (۲-۱۸)$$

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

α:	۰/۵	۰/۶۷	۰/۷۵	۰/۸۴	۰/۹۰	۰/۹۵
$\omega^{(\alpha)}$:	۱	۳/۰۲	۵/۵۲	۱۲/۸۱	۲۸/۱۱	۸۰/۰۴

(۲-۱۹)

این مقادیر را می توان با مقادیری که از محاسبات کلسترومین به دست آمده مقایسه کرد. مثلاً $\hat{\omega}(0/67) \doteq 3/9$ ، $\hat{\omega}(0/75) \doteq 6/31$ ، $\hat{\omega}(0/84) \doteq 14/3$ ، $\hat{\omega}(0/90) \doteq 24/8$.

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

$$\phi^{(0/75)}(z) = \begin{cases} \phi(z)/c & z \le z^{(0/75)} = 0/674 \\ 5/52\phi(z)/c & z > 0/674 \end{cases} \qquad (۲-۲۰)$$

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

$$\int_{-\infty}^{+\infty} z\phi^{(0/75)}(z)dz = 0/675 \qquad (۲-۲۱)$$

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

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

برای تمام مقادیر α ، $۱ > \alpha > ۰$ این رابطه برقرار است. بنابراین $\phi^{(\alpha)}$ شبیه به خانوادهٔ مکانی[۲۲] عمل می‌کند. این نتیجه مخصوص توزیع نرمال است.

یادآوری ۲ــ۴. تابع زیان‌مربع خطای نامتقارن (۲ــ۴) خاصیت مهم ناوردایی را نتیجه می‌دهد و اگر بردار مشاهدات $(y_1, \cdots, y_n)$ به $(cy_1, \cdots, cy_n)$ تغییر کند آنگاه پاسخ رابطهٔ (۲ــ۵) از β_{RP_ω} به $c\beta_{RP_\omega}$ تغییر خواهد کرد. بنابر قضیهٔ ۱ که توسط نیوی وپاول[۸] گفته شد، برآورد β_{RP_ω} دارای خاصیت ناوردایی درمکان و مقیاس[۲۳] است و داریم :

$$y \longrightarrow cy + d1 \Longrightarrow \beta_{RP_\omega} \longrightarrow c\beta_{RP_\omega} + (d, 0, 0, \cdots, 0)'$$

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

ناوردایی مقیاس فقط منحصربه تابع زیان‌مربع خطای نامتقارن نیست، بلکه هرتابع زیان توانی [۲۴] به صورت :

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

ناوردایی مقیاس را برای برآوردگر متناظر (یعنی β_ω) نتیجه می‌دهد. از رابطهٔ فوق به سادگی $r_\omega(c\lambda) = |c|^p r_\omega(\lambda)$ نتیجه می‌شود که در میان تمام حالتهای $p > 0$، حالت $(p = 2)$ دارای مزیت محاسباتی می‌باشد.

یادآوری ۲ــ۵. خاصیت سازگاری برآوردگرهای صدکهای رگرسیونی توسط نیوی وپاول (۱۹۸۷)[۸] بیان شده است. آنها در لم‌ــA نتایج کلی سازگاری در همگرایی صدکهای رگرسیونی را به متناظرشان در جمعیت بیان می کنند.

۳ یک مثال تجربی

اطلاعات مربوط به قندخون ناشتایی (FBS)[۲۵] و هموگلوبین $A1$ $(HbA1)$[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 $n = ۲۱۶$ نفر از بیمارانی است که زیر نظر قرار گرفته‌اند.

در شکل (۳ــ۱) نمودار پراکندگی (x_i, y_i) را همراه با خطوط صدکهای رگرسیونی که براساس روش (۱ــ۱) به دست آمده، مشاهده می کنیم که برای مقادیر $\alpha = ۰/۰۵, ۰/۱۰, ۰/۳۳, ۰/۵۰, ۰/۶۷, ۰/۹۰, ۰/۹۵$ رسم شده است.

در شکل (۳ــ۲) نمودار پراکندگی (x_i, y_i) همراه با خطوط صدکهای رگرسیونی برای $\omega = ۰/۰۲$ ، $\omega = ۰/۰۵$ ، $\omega = ۱$ ، $\omega = ۲۵$ ، $\omega = ۶۷$ رسم شده است.

با توجه به شکل (۳ــ۱) 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 .

درجدول (۳ــ۱) برآورد صدکهای رگرسیونی را برای ω های مختلف به دست آورده و رابطهٔ آن را با $p(\omega) = \alpha$ نشان داده‌ایم. در این جدول برای مقادیر مختلف ω ضرایب رگرسیونی و α های متناظر آن محاسبه شده و همانطورکه مشاهده می‌شود ضریب قندخون ناشتایی یک روند مثبت دارد بدین معنی که با افزایش آن مقدار هموگلوبین $A1$ افزایش پیدا می‌کند.

---

[۲۲] Location Family

[۲۳] Scale and location invariance

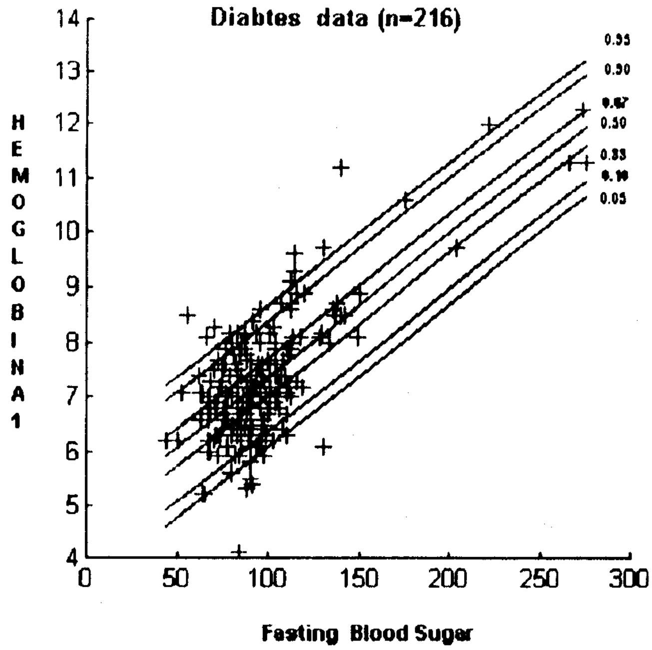
[۲۴] Power loss function

[۲۵] Fasting Blood Sugar

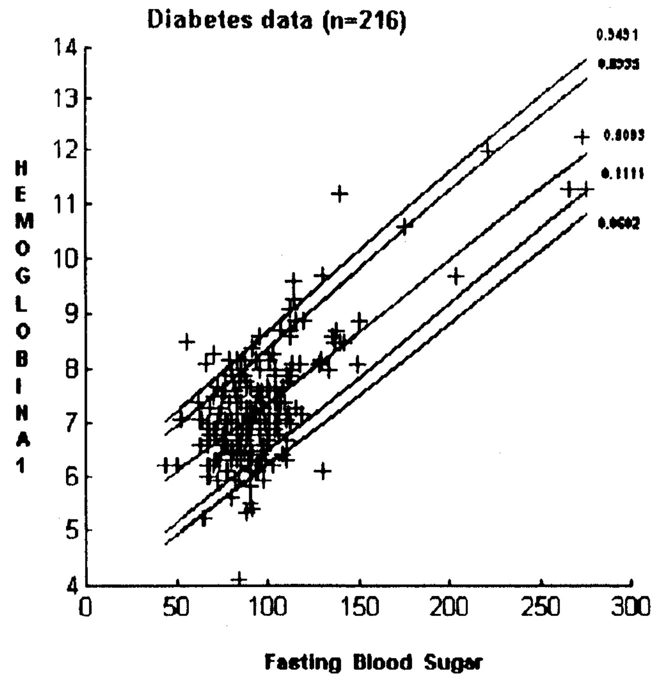
[۲۶] Hemoglobin A1

[۲۷] Diabetes

[۲۸] این اطلاعات توسط آقای دکترامینی رئیس مرکز تحقیقات غددومتابولیسم بیمارستان امین‌اصفهان جمع آوری شده است.



شکل ۳–۱ . این شکل مربوط به بیماران دیابتی‌است که‌اطلاعات مربوط به قندخون ناشتایی (FBS) و هموگلوبین $A1$ $(HbA1)$ آنهابرای زوجهای مرتب (x_i, y_i) در مقابل هم، و نیز منحنی حداقل مربعات معمولی (OLS یا $\alpha = 0.50$) برازانده شده به صورت $y_i = b_0 + b_1 x_i + e_i$ همراه با صدکهایی که از روش (۱–۱)برای مقادیر $\alpha = ۰/۰۵, ۰/۱۰, ۰/۳۳, ۰/۵۰, ۰/۶۷, ۰/۹۰, ۰/۹۵$ رسم شده است.



شکل ۳–۲. این شکل مربوط به بیماران دیابتی‌است که اطلاعات مربوط به قندخون ناشتایی (FBS) و هموگلوبین $A1$ $(HbA1)$ آنهابرای زوجهای مرتب (x_i, y_i) در مقابل هم ، و نیز صدکهای رگرسیونی برای $\omega = ۰/۰۲$ ، $\omega = ۰/۰۵$ ، $\omega = ۱$ ، $\omega = ۲۵$ ، $\omega = ۶۷$ رسم شده است .

جدول ۳ـ۱ــ تحلیل صدکهای رگرسیونی برای بیماران دیابتی

w	β_1	β_2	$p(w)$
۰/۰۲	۳/۵۸۴۵	۰/۰۲۶۳	۰/۰۶۰۲
۰/۰۵	۳/۷۷۸۴	۰/۰۲۷۱	۰/۱۱۱۱
۰/۰۸	۳/۹۵۲۶	۰/۰۲۶۸	۰/۱۴۸۱
۰/۱	۴/۰۳۴۷	۰/۰۲۶۶	۰/۱۷۵۹
۰/۱۱	۴/۰۶۹۷	۰/۰۲۶۶	۰/۱۸۰۶
۰/۱۲	۴/۱۰۱۹	۰/۰۲۶۵	۰/۱۸۹۸
۰/۱۵	۴/۱۸۲۷	۰/۰۲۶۴	۰/۲۱۳۰
۰/۳	۴/۴۱۸۲	۰/۰۲۶۰	۰/۳۲۸۷
۰/۵	۴/۵۷۳۴	۰/۰۲۶۰	۰/۴۰۲۸
۱	۴/۷۷۱۲	۰/۰۲۶۱	۰/۵۰۹۳
۱/۵	۴/۸۸۱۹	۰/۰۲۶۲	۰/۵۶۹۴
۲	۴/۹۵۷۴	۰/۰۲۶۴	۰/۶۳۴۳
۲/۲	۴/۹۸۲۴	۰/۰۲۶۴	۰/۶۴۳۵
۹	۵/۲۸۲۶	۰/۰۲۸۱	۰/۷۹۱۷
۱۰	۵/۳۰۲۱	۰/۰۲۸۲	۰/۸۱۰۲
۱۲	۵/۳۴۰۲	۰/۰۲۸۳	۰/۸۲۴۱
۱۳	۵/۳۵۷۱	۰/۰۲۸۴	۰/۸۲۴۱
۲۵	۵/۵۲۶۲	۰/۰۲۸۶	۰/۸۹۳۵
۲۷	۵/۵۴۹۳	۰/۰۲۸۶	۰/۸۹۳۵
۲۹	۵/۵۶۹۹	۰/۰۲۸۶	۰/۸۹۸۱
۳۰	۵/۵۷۸۶	۰/۰۲۸۶	۰/۹۰۲۸
۵۰	۵/۷۰۴۰	۰/۰۲۸۹	۰/۹۴۴۴
۵۵	۵/۷۲۷۹	۰/۰۲۹۰	۰/۹۴۴۴
۶۰	۵/۷۵۱۱	۰/۰۲۹۰	۰/۹۴۹۱
۶۵	۵/۷۷۲۰	۰/۰۲۹۱	۰/۹۴۹۱

مراجع

[1] D.Aigner, T.Amemiya, and D.Poirier,"On the estimation of production fronters: maximum likelihod estimation of the parameters of a discountinus 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، احمدرضا: «صدکهای رگرسیونی با استفاده‌از تابع زیان مربع نامتقارن خطا» پایان نامه (۱۳۷۷)دانشگاه صنعتی اصفهان

[۱۰] کنوور: «آمارناپارامتری کاربردی » چاپ اول (۱۳۷۲)مرکزنشر دانشگاهی‌تهران. مترجم: هاشمی‌پرست، سیدمقتدی