

## صدکهای رگرسیونی با استفاده از تابع زیان مربع خطای نامتقارن

احمد رضا جمشیدیان      سروش علیمردی<sup>۱</sup>

### چکیده

خروجی اولیه در تحلیل رگرسیونی معمولی برآورد تابع رگرسیونی است که نقاط وسط انبوه نقاط را نمایش می دهد. 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 نمونه ای 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 کرده ایم.

### ۱ مقدمه

داده هائی که برای انجام یک تحلیل رگرسیونی به کار می روند به فرم انبوهی از نقاط در فضای اقلیدسی به صورت زیر هستند:

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

افرون<sup>۲</sup> (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  $Y$  به عنوان تابعی از  $X$  برآورد می شوند.

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

شکل (۱-۱) یک رگرسیون ساده را نشان می دهد. «ساده» بدین مفهوم که بردار  $x_i$  تک بعدی است و داده ها را می توان به وسیله نمودار پراکندگی، به صورت زوجهای  $(x_i, y_i)$  نشان داد. این شکل مربوط به یک آزمایش کلینیکی در استفاده از

<sup>۱</sup> - احمد رضا جمشیدیان، سروش علیمردی، گروه آمار دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان  
<sup>۲</sup> B. Efron

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 آمین تا ۹۰ آمین صدک برآورد می‌شود تا یک کاهش کلسترول خون مابین ۶- تا ۵۴ واحد در مجموعه کل کلسترولهای شمارش شده باشد.

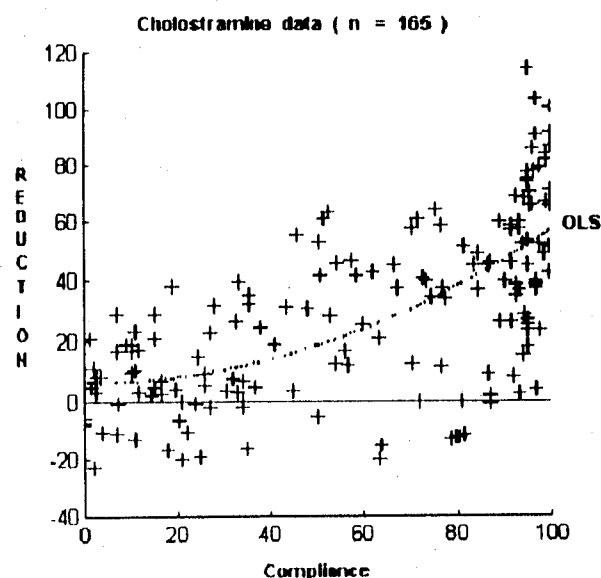
روش متداول برای نشان دادن صدکهای رگرسیونی بدین صورت است. فرض کنید میزان خطا در مدل رگرسیون همگن و نرمال (یعنی  $e_i \sim N(0, \sigma^2)$  برای  $i = 1, 2, 3, \dots, n$ ) باشد، برآورد  $\sigma^2$  را بصورت  $\hat{\sigma}^2$  نشان می‌دهیم، آنگاه برآورد  $\alpha$  ۱۰۰ آمین صدک رگرسیونی به صورت:

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

به دست می‌آید [۹] که  $\hat{\mu}(x)$  برآورد رگرسیون حداقل مربعات معمولی و  $Z^{(\alpha)}$  هم  $\alpha$  ۱۰۰ آمین صدک از تابع چگالی نرمال استاندارد  $Z^{(\alpha)} = \Phi^{-1}(\alpha)$  است. اگر تمام این خطوط را برای  $\alpha$  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<sup>۸</sup> (۱۹۷۸) [۶] می‌باشد. آنها از قدرمطلق زیان نامتقارن<sup>۹</sup> بیشتر از مربع زیان نامتقارن<sup>۱۰</sup> استفاده کرده، در این بین چندکهای رگرسیونی<sup>۱۱</sup> را تعریف کردند. پاول و نیوی<sup>۱۲</sup> (۱۹۸۷) [۸] کاری را که توسط آیگنر، آمییا و پوپیر<sup>۱۳</sup> (۱۹۷۶) [۱] انجام شده بود، ادامه دادند و از مربع

داروی کلسترومین<sup>۲</sup> و تأثیر آن بر پایین آوردن کلسترول خون<sup>۴</sup> است. بنابراین  $y_i$  میزان کاهش<sup>۵</sup> خون و  $x_i$  درصد بسته‌های کلسترومین مصرفی است که بیمار<sup>۶</sup> ام قبول کرده است. منحنی حداقل مربعات معمولی برازنده شده، یک مدل رگرسیونی درجه دو به صورت  $y_i = b_0 + b_1 x_i + b_2 x_i^2 + e_i$  است<sup>۷</sup>.



شکل ۱-۱: این شکل مربوط به آزمایش کلینیکی در استفاده از داروی کلسترومین و تأثیر آن بر پایین آوردن کلسترول خون، همراه با منحنی حداقل مربعات معمولی برازنده 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 وسیله روش حداقل مربعات نامتقارن، که در اینجا

۷- این اطلاعات مربوط به بیماران تحت کنترل بخش استانفورد می‌باشد که در *LipidResearchClinicsProgram* (۱۹۸۴) [۷] توسط Drs.D.FeldmanandJ.Farquhar از دانشکده داروسازی استانفورد جمع آوری شده است.

- ۲-Cholestyramine
- ۴-Blood Cholesterol
- ۵-Reduction
- ۶-Compliance
- 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

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:

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

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

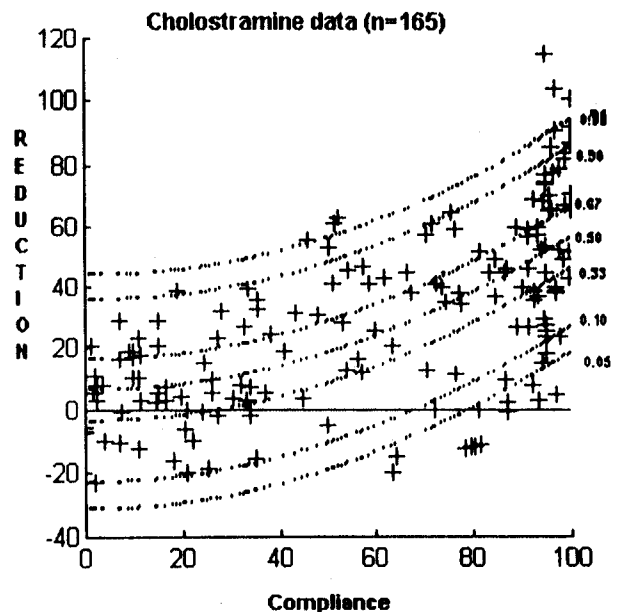
برای برآورد صدکهای رگرسیونی پارامتر  $\beta$ ، ابتدا تابع زیان مربع خطای نامتقارن را برای هر  $\omega > 0$  به فرم:

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

تعریف کرده و سپس معادله:

$$S_\omega(\beta) = \sum_{i=1}^n r_\omega(y_i - x_i' \beta) \quad (5-2)$$

خطای نامتقارن آنچنان که ما در اینجا از آن استفاده خواهیم کرد، استفاده کردند و نتایج منحنی را انتظارک<sup>۱۲</sup> 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<sup>۱۵</sup> [۵] پارامترهای رگرسیونی بوده‌اند.



شکل ۱-۲: این شکل مربوط به صدکهای رگرسیونی متداول برای آزمایش کلینیکی در استفاده از داروی کلسترومین و تاثیر آن بر پایین آوردن کلسترول خون است و برای  $\alpha = 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95$  صدکهای رگرسیونی آن رسم شده است.

در اینجا صدکهای رگرسیونی که اولین بار توسط افرون (۱۹۹۱) [۳] ارایه شد جالبترین وی صدکهای نمونه‌ای را که در توصیف اطلاعات نمونه‌ای مفیدند، براساس چندکهای رگرسیونی تعریف شده توسط کانکر و باست [۶] به رگرسیون

Expectile—<sup>۱۲</sup>  
Robustify—<sup>۱۵</sup>

$\hat{L}^{(\alpha)}$  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  $U(b)$  دارای عناصری به صورت زیر است:

$$u_i(b) = y_i - x_i' b \quad i = 1, 2, \dots, n \quad (6-2)$$

در این حالت معادله  $(2-2)$  به صورت زیر درمی‌آید و صدکهای رگرسیونی برحسب  $b$  از مینیم 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  $\beta_{RPw}$  منحصرأ 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 می‌کنیم:

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

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

$$r_w(\lambda) = W(\lambda)\lambda^2$$

می‌شود و با مشتق‌گیری از این رابطه داریم:

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

که یک تابع پیوسته از  $\lambda$  می‌باشد. پس می‌توان نتیجه گرفت که بردار گرادیان  $S'_w(b)$  یک تابع پیوسته از  $b$  است. بنابراین:

را برحسب  $\beta$  مینیم می‌کنیم و حاصل را با  $\beta_{RPw}$  نشان می‌دهیم.

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

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

یادآوری ۲-۲. به ازای تمام مقادیر  $\omega > 0$ ، آبرصفحه  $\hat{L}_\omega$  بعدی  $\hat{L}_\omega$  را به وسیله  $\beta_{RPw}$  در فضای  $P+1$  بعدی  $(x_i, y)$  به صورت زیر تعریف می‌کنیم:

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

که این را «صفحه  $\omega$  - رگرسیونی»<sup>۱۶</sup> می‌نامیم. این صفحه به طور یکنواخت از پایین تا بالای انبوه داده‌ها با افزایش  $\omega$  از صفر تا بینهایت حرکت می‌کند.

$P(\omega)$  را نسبت تعداد نقاط  $(x_i, y_i)$  به کل داده‌ها که زیر صفحه  $\hat{L}_\omega$  قرار می‌گیرند (یا به عبارت دیگر، نسبت تعداد نقاطی که  $u_i(\beta_{RPw}) \leq 0$ )،  $(2-6)$  به تعداد کل داده‌ها تعریف می‌کنیم. برای  $P(\omega) = \alpha$  (که  $\alpha$  مقادیری بین صفر و یک رami گیرد) داریم:

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

به عبارت دیگر برای مقدار مفروضی از  $\alpha$ ،  $0 < \alpha < 1$ ، بردار  $\beta_{RP}^{(\alpha)}$  همان بردار  $\beta_{RPw}$  برای مقادیری از  $\omega$  است که بیانگر  $\alpha$  درصد از داده‌هایی است که زیر صفحه  $\hat{L}_\omega$  قرار دارند. بنابراین سطوح متناظر با  $\beta_{RP}^{(\alpha)}$  را به صورت:

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

تعریف می‌کنیم که در واقع  $\alpha 100$  امین صدک رگرسیونی است. به عبارت دیگر صفحات مختلف  $\hat{L}_\omega$  را می‌توان به صورت

Hyperplane—<sup>۱۶</sup>  
 $\omega$ -regression plane—<sup>۱۷</sup>  
 Gradient vector—<sup>۱۸</sup>

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

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

که:

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

اگر هیچیک از مقادیر  $u_i(b)$  صفر نباشد، آنگاه مشتق دوم حتماً پیوسته است.

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

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

که برای حل معادله یاد شده احتیاج به روشهای تکراری<sup>۱۹</sup> داریم. از این معادله می‌توان نتیجه گرفت:

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

و اگر پاسخ رابطه 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 وزنه‌های  $W(u_i(b))$  ثابت باشند، پاسخ گاوس مارکف<sup>۲۰</sup> برای انتخاب  $\beta$  در جهت مینیم کردن مجموع مربعات:

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

$\beta(\omega, b)$  است، آنگاه:

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

و این مقدار با توجه به معادله (۱۰-۲) برابر با صفر است. بنابراین با توجه به خاصیت مانایی<sup>۲۱</sup> داریم:

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

Iterative methods—<sup>۱۹</sup>

Gauss Markov—<sup>۲۰</sup>

Stationary—<sup>۲۱</sup>

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

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

از طرفی:

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

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

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

مستقل وجود ندارد، (یعنی برای هر  $i$ ،  $x_i = 1$ )، وقتی  $n \rightarrow \infty$  بافت‌نگار  $y_1, \dots, y_n$  به سمت توزیع  $N(0, 1)$  میل می‌کند. در این وضعیت  $\beta_{RP\omega}$  را با  $\beta_\omega$  نشان می‌دهیم.  $\omega^{(\alpha)}$  بیانگر مقدار  $\omega$  است که  $\beta_\omega$  برابر با  $Z^{(\alpha)}$  باشد.  $Z^{(\alpha)} = \Phi^{-1}(\alpha)$  نقطه  $\alpha$  ۱۰۰ 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  $(\omega^{(\alpha)} = \frac{1}{\omega(1-\alpha)})$ . که با توجه به این فرمول مقادیر زیر را به دست می‌آوریم:

$$\begin{array}{cccccc} \alpha: & 0/5 & 0/67 & 0/75 & 0/84 & 0/90 & 0/95 \\ \omega^{(\alpha)}: & 1 & 3/02 & 5/52 & 12/81 & 28/11 & 80/04 \end{array} \quad (19-2)$$

این مقادیر را می‌توان با مقادیری که از محاسبات کلسترومین به دست آمده مقایسه کرد. مثلاً  $\hat{\omega}(0/75) \doteq 6/31$ ،  $\hat{\omega}(0/67) \doteq 3/9$ ،  $\hat{\omega}(0/90) \doteq 24/8$ ،  $\hat{\omega}(0/84) \doteq 14/3$  با تغییر دادن تابع چگالی نرمال استاندارد  $\phi(z)$  به صورت:

$$\phi^{(0/75)}(z) = \begin{cases} \phi(z)/c & z \leq z^{(0/75)} = 0/674 \\ 5/52\phi(z)/c & z > 0/674 \end{cases} \quad (20-2)$$

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

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

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  $\delta(\omega, b)$  را تابعی از  $\omega$  با مقدار ثابت  $b$  در نظر می‌گیریم. آنگاه خواهیم داشت:

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

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

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

که  $|U(b)|$  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  $\omega$  می‌توان پاسخ را در یک همسایگی  $\omega$  به فرم  $\omega + \Delta\omega$  به صورت زیر تقریب زد:

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

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.

$$\beta_{RP(\omega+\Delta\omega)}^{(0)} = \beta_{RP} + \frac{\partial \beta_{RP\omega}}{\partial \omega} \Delta\omega$$

&

$$\beta_{RP(\omega+\Delta\omega)}^{(1)} = \beta(\omega + \Delta\omega, \beta_{(\omega+\Delta\omega)}^{(0)}) \quad (17-2)$$

هرچه  $\Delta\omega$  کوچکتر باشد،  $\beta_{RP(\omega+\Delta\omega)}^{(1)}$  تقریب خوبی برای  $\beta_{RP(\omega+\Delta\omega)}$  می‌باشد.  $\omega$  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:

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

که در واقع برآورد مینیمم مربعات خطاهاست.

یادآوری ۲-۳. یک ایده کلی در مورد رابطه بین  $\omega$  و  $\alpha$ ، حالت خاص زیر است. در این حالت فرض می‌کنیم متغیر

### ۳ یک مثال تجربی

اطلاعات مربوط به قندخون ناشتایی (FBS)<sup>۲۵</sup> و هموگلوبین A<sub>1</sub> (HbA<sub>1</sub>)<sup>۲۶</sup> را به عنوان یک پارامتر در خون که نشان دهنده وضعیت بیماری دیابت<sup>۲۷</sup> است، در نظرمی گیریم<sup>۲۸</sup>.  $x$  (قندخون ناشتایی) به عنوان متغیر مستقل و  $y$  (هموگلوبین A<sub>1</sub>) به عنوان متغیر وابسته برای  $n = ۲۱۶$  نفر از بیماران است که زیر نظر قرار گرفته اند.

در شکل (۱-۳) نمودار پراکندگی  $(x_i, y_i)$  را همراه با خطوط صدکهای رگرسیونی که براساس روش (۱-۱) 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 (۲-۳) نمودار پراکندگی  $(x_i, y_i)$  همراه با خطوط صدکهای رگرسیونی برای  $\omega = ۱, \omega = ۰/۰۵, \omega = ۰/۰۲$  رسم شده است.  $\omega = ۶۷, \omega = ۲۵$

با توجه به شکل (۱-۳) مشاهده می شود اگر آزمایش قندخون ناشتایی فردی، با تمرکز ۹۰٪ برابر با ۱۰۰ باشد انتظار می رود هموگلوبین A<sub>1</sub> آن حدود ۶/۲ تا ۸/۷ باشد.

در جدول (۱-۳) برآورد صدکهای رگرسیونی را برای  $\omega$  های مختلف به دست آورده و رابطه آن را با  $p(\omega) = \alpha$  نشان داده ایم. در این جدول برای مقادیر مختلف  $\omega$  ضرایب رگرسیونی و  $\alpha$  های متناظر آن محاسبه شده و همانطور که مشاهده می شود ضریب قندخون ناشتایی یک روند مثبت دارد بدین معنی که با افزایش آن مقدار هموگلوبین A<sub>1</sub> افزایش پیدا می کند.

برای تمام مقادیر  $\alpha, ۰ < \alpha < ۱$  این رابطه برقرار است. بنابراین  $\phi(\alpha)$  شبیه به خانواده مکانی<sup>۲۲</sup> عمل می کند. این نتیجه مخصوص توزیع نرمال است.

یادآوری ۲-۴. تابع زبان مربع خطای نامتقارن (۲-۴) خاصیت مهم نوردایی را نتیجه می دهد و اگر بردار مشاهدات  $(y_1, \dots, y_n)$  به  $(cy_1, \dots, cy_n)$  تغییر کند آنگاه پاسخ رابطه (۲-۵) از  $\beta_{RP\omega}$  به  $c\beta_{RP\omega}$  تغییر خواهد کرد. بنابر قضیه ۱ که توسط نیوی وپاول [۸] گفته شد، برآورد  $\beta_{RP\omega}$  دارای خاصیت نوردایی در مکان و مقیاس<sup>۲۳</sup> است و داریم:

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

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

نوردایی مقیاس فقط منحصربه تابع زبان مربع خطای نامتقارن نیست، بلکه هرتابع زبان توانی<sup>۲۴</sup> به صورت:

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

نوردایی مقیاس را برای برآوردگر متناظر (یعنی  $\beta_\omega$ ) نتیجه می دهد. از رابطه فوق به سادگی  $r_\omega(c\lambda) = |c|^p r_\omega(\lambda)$  نتیجه می شود که در میان تمام حالت های  $p > 0$ ، حالت  $(p = ۲)$  دارای مزیت محاسباتی می باشد.

یادآوری ۲-۵. خاصیت سازگاری برآوردگرهای صدکهای رگرسیونی توسط نیوی وپاول (۱۹۸۷) [۸] بیان شده است. آنها در لم A- نتایج کلی سازگاری در همگرایی صدکهای رگرسیونی را به متناظرشان در جمعیت بیان می کنند.

۲۲- Location Family

۲۳- Scale and location invariance

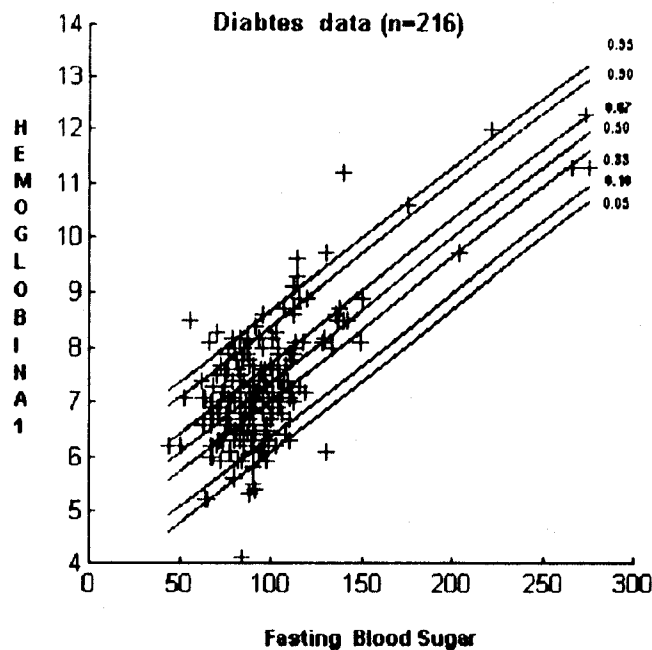
۲۴- Power loss function

۲۵- Fasting Blood Sugar

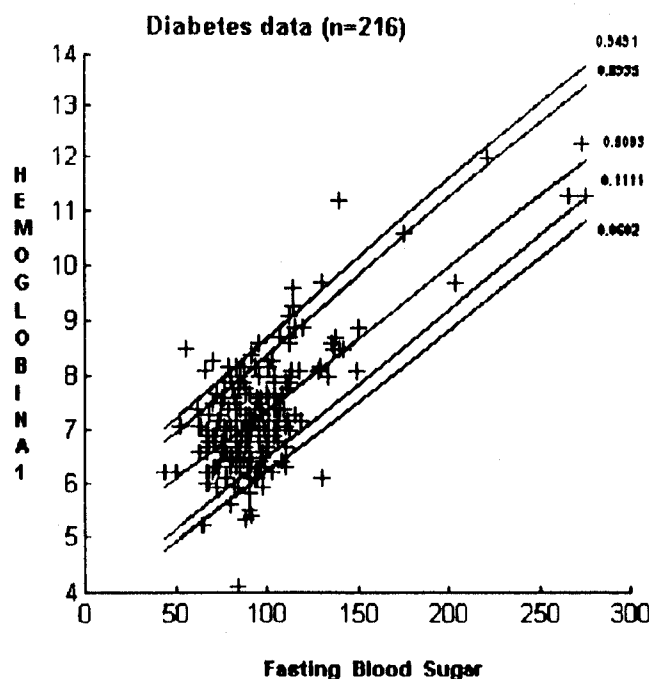
۲۶- Hemoglobin A<sub>1</sub>

۲۷- Diabetes

۲۸- این اطلاعات توسط آقای دکترامینی 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 ناشتایی ( $FBS$ ) و هموگلوبین  $A_1$  ( $HbA_1$ ) آنها برای زوجهای مرتب  $(x_i, y_i)$  در مقابل هم، و نیز منحنی حداقل مربعات معمولی (OLS یا  $\alpha = 0.50$ ) برازنده شده به صورت  $y_i = b_0 + b_1 x_i + e_i$  همراه با صدکهایی که از روش (۱-۱) برای مقادیر  $\alpha = 0.05, 0.10, 0.27, 0.50, 0.67, 0.90, 0.95$  رسم شده است.



شکل ۲-۳. این شکل مربوط به بیماران دیابتی است که اطلاعات مربوط به قندخون ناشتایی ( $FBS$ ) و هموگلوبین  $A_1$  ( $HbA_1$ ) آنها برای زوجهای مرتب  $(x_i, y_i)$  در مقابل هم، و نیز صدکهای رگرسیونی برای  $\omega = 1, \omega = 0.05, \omega = 0.02$  رسم شده است.  $\omega = 67, \omega = 25$

جدول ۳-۱- تحلیل صدکهای رگرسیونی برای بیماران دیابتی

| $w$  | $(\beta_1, \beta_2)$ | $p(w)$ |
|------|----------------------|--------|
| 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 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، احمدرضا: «صدکهای رگرسیونی با استفاده از تابع زیان مربع نامتقارن خطا» پایان نامه (۱۳۷۷) دانشگاه صنعتی اصفهان
- [۱۰] کنوور: «آمارناپارامتری کاربردی» چاپ اول (۱۳۷۲) مرکز نشر دانشگاهی تهران. مترجم: هاشمی پرست، سیدمقتدی