

## مقایسه مدل های حاشیه ای برای تحلیل داده های طولی با پاسخ های دودویی

مجتبی گنجعلی<sup>۱</sup>، امید امدادی فر<sup>۲</sup>

دانشگاه شهید بهشتی، دانشکده علوم ریاضی، گروه آمار

(دریافت: ۸۱/۴/۹، پذیرش: ۸۱/۸/۱۴)

### چکیده

مدل حاشیه ای (Marginal model) یکی از رهیافت هائی است که برای تحلیل داده های طولی (Longitudinal data) بکار می رود. در این مدل، همبستگی موجود بین پاسخ ها برای هر آزمودنی (Individual) به عنوان پارامتر در نظر گرفته می شود تا تحلیل های معتبری را نتیجه دهد. مدل های حاشیه ای مختلفی که برای تحلیل داده های طولی با پاسخ های دودویی ذکر شده اند از جمله مدل های حاشیه ای با لگ نسبت بخت حاشیه ای (Marginal odds ratio)، لگ نسبت بخت شرطی (Conditional odds ratio)، لگ نسبت وابستگی (Dependence ratio)، مدل پروبیت چند متغیره (Multivariate probit model) و روش معادلات برآوردگر تعمیم یافته [Generalized Estimating Equations (GEE)] مرور و مقایسه شده اند. مانده هایی برای بررسی نیکویی برازش این مدل ها معرفی گردیده است. همچنین مدل های ذکر شده را درمنالی کاربردی برازش داده ایم.

**واژه های کلیدی:** مطالعات طولی، مدل های حاشیه ای، معادلات برآوردگر تعمیم یافته، نسبت وابستگی، نسبت بخت حاشیه ای، نسبت بخت شرطی، مدل پروبیت چند متغیره

## ۱- مقدمه

داده‌های طولی داده‌هایی هستند که در آنها پاسخ‌ها برای هر آزمودنی در طول زمان تکرار شده‌اند و چون اغلب بین پاسخ‌ها برای هر آزمودنی خاص همبستگی وجود دارد نیاز به روش‌های ویژه‌ای برای تحلیل وجود دارد تا بتوان استنباط‌های معتبری به دست آورد. چندین رهیافت برای تحلیل چنین داده‌هایی بکار می‌روند که از جمله آنها می‌توان مدل‌های اثرات تصادفی (Random effects models)، مدل‌های انتقالی (Transition models) و مدل‌های حاشیه‌ای را نام برد (Diggel, et al., 1994).

در بخش دوم مدل حاشیه‌ای را تعریف کرده، سپس انواع مدل‌های حاشیه‌ای را که برای داده‌های طولی دودویی بکار می‌روند در بخش‌های ۳ تا ۶ مرور کرده‌ایم. در بخش ۷ مقایسه‌ای کلی بین مدل‌های حاشیه‌ای ارائه شده، انجام داده‌ایم. از آنجا که در داده‌های طولی دودویی مانده‌هایی که بتوانند با در نظر گرفتن همبستگی پاسخ‌های یک آزمودنی خاص نیکویی برازش مدل را انجام دهند هنوز معرفی نشده‌است، مانده‌هایی برای مدل‌های ذکر شده در بخش ۸ معرفی کرده‌ایم. سپس مدل‌های بیان شده را در بخش ۹ بر روی داده‌های آسم (Rotnitzky & Wypij, 1994) برازش داده‌ایم.

## ۲- مدل‌های حاشیه‌ای

مدل‌های حاشیه‌ای مدل‌هایی هستند که در آنها تاثیر متغیرهای تبیینی بر پاسخ‌ها به طور مجزا از همبستگی بین پاسخ‌ها برای یک آزمودنی معین مدل‌بندی می‌شود. در این نوع مدل‌ها، امید حاشیه‌ای  $E(Y_{ij})$  به صورت تابعی از متغیرهای تبیینی مدل‌بندی می‌شود. امید حاشیه‌ای همان پارامتری است که در مطالعه مقطعی نیز مدل‌بندی می‌شود (Agresti, 1990). در یک مدل حاشیه‌ای امید حاشیه‌ای پاسخ  $\mu_{ij} = E(Y_{ij})$  به متغیرهای تبیینی  $x_{ij}$  بصورت  $h(\mu_{ij}) = x'_{ij}\beta$  وابسته است؛ که در آن  $h$  تابع پیوند (Link function) شناخته شده‌ای است. به عنوان مثال  $h$  برای پاسخ‌های دودویی، لوچیت می‌باشد. همچنین واریانس حاشیه‌ای به میانگین حاشیه‌ای به صورت  $\text{var}(Y_{ij}) = v(\mu_{ij})\phi$  وابسته است که  $v$  تابع واریانس شناخته شده و  $\phi$  پارامتر مقیاسی است که ممکن است، علاوه بر دیگر پارامترها نیاز به برآورد داشته باشد و همبستگی  $Y_{ij}$  و  $Y_{ik}$  به صورت تابعی از میانگین حاشیه‌ای آنها و پارامترهای اضافی  $\alpha$  - است یعنی:

$$\text{Corr}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \alpha)$$

که  $\rho(\cdot)$  تابع شناخته شده‌ای از  $\mu_i$  ها است. ضرایب رگرسیونی حاشیه‌ای  $\beta$  - دارای تفسیری مشابه ضرایب رگرسیونی در تحلیل مقطعی هستند. می‌توان گفت مدل‌های حاشیه‌ای برای پاسخهای وابسته تعمیمی از مدل‌های خطی تعمیم یافته (Generalized linear models) برای پاسخهای مستقل می‌باشند.

### ۳- رهیافت بهادر (Bahadur)

بهادر (۱۹۶۱) توزیع توام  $(Y_{i1}, \dots, Y_{im})$  که در آن  $Y_{ij}$  برای  $j=1, \dots, m$  و  $i=1, \dots, n$  پاسخهای دودویی هستند را به صورت زیر در نظر می‌گیرد.

$$\prod_{j=1}^m \mu_{ij}^{y_{ij}} (1 - \mu_{ij})^{1-y_{ij}} \left( 1 + \sum_{j < k} \rho_{ijk} e_{ij} e_{ik} + \sum_{j < k < l} \rho_{ijkl} e_{ij} e_{ik} e_{il} + \dots + \rho_{i12 \dots n} e_{i1} e_{i2} \dots e_{in} \right)$$

که در آن:

$$\rho_{i1 \dots n} = E(e_{i1} e_{i2} \dots e_{in}), \dots, \rho_{ijk} = E(e_{ij} e_{ik}), \quad e_{ij} = \frac{y_{ij} - \mu_{ij}}{\{\mu_{ij}(1 - \mu_{ij})\}^{1/2}}$$

بنا بر این توزیع توام پاسخها بوسیله فرضیهایی درباره  $n-1$  همبستگی‌های حاشیه‌ای

$$\rho_i = (\rho_{i12}, \dots, \rho_{im-1n}, \rho_{i12 \dots n})'$$

مشخص می‌شود.

نمایش بهادر در بالا برای درجه‌های متفاوتی از وابستگی در میان  $Y_{ij}$  ها قابل بیان می‌باشد. مثلا ساختار استقلال نتیجه می‌دهد که تمام همبستگی‌های حاشیه‌ای دودویی و مراتب بالاتر صفر است. اما همبستگی‌های حاشیه‌ای (که در مدل بهادر تعدادشان با افزایش  $n$  بسیار زیاد می‌شود) به صورت زیر بوسیله احتمال‌های حاشیه‌ای (میانگین‌های حاشیه‌ای که در پاسخهای دودویی احتمال‌های پیروزی حاشیه‌ای اند) محدود می‌شوند.

$$A \leq \text{Corr}(Y_{ij}, Y_{ik}) \leq B \quad (1)$$

که در آن:

$$A = \max \left\{ - \left( \frac{\mu_{ij} \mu_{ik}}{(1 - \mu_{ij})(1 - \mu_{ik})} \right)^{1/2}, - \left( \frac{(1 - \mu_{ij})(1 - \mu_{ik})}{\mu_{ij} \mu_{ik}} \right)^{1/2} \right\}$$

$$B = \min \left\{ \left( \frac{\mu_{ij}(1-\mu_{ik})}{(1-\mu_{ij})\mu_{ik}} \right)^{1/2}, \left( \frac{(1-\mu_{ij})\mu_{ik}}{\mu_{ij}(1-\mu_{ik})} \right)^{1/2} \right\}$$

به همین دلیل بسیاری از محققین، معیارهایی برای همبستگی در نظر می‌گیرند که موجب می‌شود حدود همبستگی به میانگین‌ها مقید نباشند. آنها برای آزمودنی  $i$  ام  $(i = 1, \dots, m)$ ، تابع چگالی توأم پاسخ‌ها را به صورت زیر در نظر می‌گیرند:

$$f(y_i, \Psi_i, \Omega_i) = \exp \{ \Psi_i' y_i + \Omega_i' W_i - A(\Psi_i, \Omega_i) \} \quad (2)$$

که در آن:

$$W_i = (y_{i1} y_{i2} \dots y_{i(n-1)} y_{in} \dots y_{i1} y_{i2} \dots y_{in})$$

برداری  $1 \times (2^n - n - 1)$  از حاصلضرب‌های دوتایی و بالاتر  $y_i$  و  $\Psi_i = (\psi_{i1}, \dots, \psi_{in})'$  و بردارهایی از پارامترهای متعارف می‌باشند.  $\Omega_i = (\omega_{i12}, \dots, \omega_{i(n-1)n}, \dots, \omega_{i12, n})'$  مقدار ثابتی است که به  $y_i$  وابسته نیست.

#### ۴- رهیافت‌های حاشیه‌ای دیگر براساس درستنمایی

در این بخش مدل‌های حاشیه‌ای براساس درستنمایی ارائه می‌شوند. در تمام این مدل‌ها، توزیع توأم پاسخ‌ها بوسیله در نظر گرفتن فرض‌هایی به طور کامل مشخص می‌شوند. یکی از این فرض‌ها درباره ارتباط پاسخ‌ها است. در این بخش رهیافت‌های حاشیه‌ای که معیارهای همبستگی آنها مقید به احتمال‌های حاشیه‌ای نیستند \_ لگ نسبت بخت شرطی، نسبت وابستگی و لگ نسبت بخت حاشیه‌ای \_ بیان می‌شوند.

در تمام این مدل‌ها ارتباط پاسخ‌ها با متغیرهای کمکی بصورت  $\logit(\mu_{ij}) = x'_{ij} \beta$  و واریانس پاسخ‌ها بصورت  $\text{var}(Y_{ij}) = \mu_{ij}(1-\mu_{ij})$  بیان می‌شود. تنها تفاوت آنها در معیار همبستگی پاسخ‌ها است. توجه داشته باشید که لوجیت  $(\logit)$  به معنای لگاریتم بخت است یعنی:

$$\logit(\mu_{ij}) = \log \left( \frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)} \right)$$

(۴-۱) مدل حاشیه‌ای با پارامترهای همبستگی به صورت توابعی از لگ نسبت بخت شرطی در این مدل‌بندی که توسط فیتز موریس و لرد پیشنهاد شده است (Fitzmaurice & Laird, 1993). در معادله ۲ همبستگی پاسخ‌ها برای هر آزمودنی براساس لگ نسبت بخت شرطی و توابعی از آنها قابل بیان است.

در معادله ۲ پارامترهای  $\Psi_i$  عباراتی از احتمال‌های شرطی به صورت:

$$\psi_{ij} = \log it \{ \Pr(y_{ij} = 1) \}$$

و پارامترهای  $\Omega_i$  (تعریف شده در معادله ۲) عباراتی از لگ نسبت بخت شرطی است یعنی:

$$\exp(\omega_{irs}) = \frac{\Pr(Y_{ir} = 1, Y_{is} = 1 | Y_{it} = 0, t \neq r, s)}{\Pr(Y_{ir} = 1, Y_{is} = 0 | Y_{it} = 0, t \neq r, s)} \times \frac{\Pr(Y_{ir} = 0, y_{is} = 0 | Y_{it} = 0, t \neq r, s)}{\Pr(Y_{ir} = 0, y_{is} = 1 | Y_{it} = 0, t \neq r, s)}$$

و باید توجه کرد که  $\mu_i$  تابعی از هر دو بردار  $\Psi_i$  و  $\Omega_i$  است.

در این مدل پارامترهای همبستگی مقید به میانگین حاشیه‌ای نیستند و فضای پارامترها

برای  $\Omega_i = (\omega_{i12}, \dots, \omega_{i12, n})'$ ،  $\mathbb{R}^{2^n - n - 1}$  (فضای اقلیدسی  $(n-1) - 2^n$  بعدی) می‌باشد. ولی این نوع پارامترها به  $n$  (تعداد پاسخ‌های تکرار شده برای هر آزمودنی) وابسته می‌باشند. یعنی می‌توان گفت این نوع مدل‌بندی برای داده‌هایی با تعداد پاسخ‌های تکرار شده متفاوت برای برخی از آزمودنی‌ها قابل کاربرد نیستند.

#### ۲-۴) مدل حاشیه‌ای با پارامترهای همبستگی به صورت توابعی از نسبت وابستگی

این مدل حاشیه‌ای توسط اخولم و دیگران، پیشنهاد شده است (Ekholm et al., 1995). در این روش استفاده از پارامتر میانگین (مولفه‌هایی که احتمال موفقیت توام تمام مراتب پاسخ‌ها را در بردارد) برای تحلیل رگرسیون پاسخ دودویی چند متغیره پیشنهاد شده است که در آن همبستگی را با استفاده از نسبت‌های وابستگی (که عبارت‌هایی از پارامتر میانگین می‌باشد) بیان می‌کنند.

اگر  $S(Y_i) = (Y_i, W_i)'$  را در نظر بگیریم که  $Y_i$  و  $W_i$  همان بردارهای تعریف شده در معادله ۲ هستند آنگاه به بردار مقادیر مورد انتظار  $S(Y_i)$ ، پارامتر میانگین  $\phi_i$  می‌گویند (Barndorff-Neilson & Cox, 1994).

یعنی:

$$\phi_i = E\{S(Y_i)\} = (\phi_{i1}, \dots, \phi_{im}, \phi_{i12}, \dots, \phi_{i(n-1)m}, \dots, \phi_{i12, n})'$$

مولفه‌های  $\phi_i$  احتمال‌های موفقیت توام حاشیه‌ای را برای اندیس‌های آن بیان می‌کند یعنی:

$$\phi_{i1} = \Pr(Y_{i1} = 1), \dots, \phi_{i12} = \Pr(Y_{i1} = 1, Y_{i2} = 1), \dots,$$

$$\phi_{i12, n} = \Pr(Y_{i1} = 1, \dots, Y_{in} = 1)$$

برای مثال پاسخ دودویی دومتغیره برای آزمودنی  $i$  ام،  $(Y_{i1}, Y_{i2})$ ، با پارامتر میانگین  $\phi_i = (\phi_{i1}, \phi_{i2}, \phi_{i12})'$  را در نظر بگیرید. نسبت وابستگی،  $\omega_{i12}$ ، و لگ نسبت وابستگی،  $\lambda_{i12}$ ، به صورت زیر تعریف می‌شود:

$$\begin{aligned} \phi_{i12} &= \Pr(Y_{i1} = 1, \dots, Y_{im} = 1) \\ \omega_{i12} &= \frac{\phi_{i12}}{\phi_{i1}\phi_{i2}}, \lambda_{i12} = \log(\omega_{i12}) \end{aligned} \quad (۳)$$

که در آن  $10 \times (\omega_{i12} - 1)$  میزان درصد بزرگی احتمال موفقیت توام  $Y_{i1}$  و  $Y_{i2}$  در مقایسه با فرض استقلال را نشان می‌دهد.

نسبت‌های وابستگی مراتب بالاتر نیز به صورت زیر تعریف می‌شود:

$$\omega_{i123} = \frac{\phi_{i123}}{\phi_{i1}\phi_{i2}\phi_{i3}}, \dots, \omega_{i123 \dots n} = \frac{\phi_{i123 \dots n}}{\phi_{i1}\phi_{i2} \dots \phi_{in}}$$

نسبت‌های وابستگی بوسیله مقایسه احتمال‌های موفقیت حاشیه‌ای با حالت استقلال تفسیر می‌شوند، اگر نسبت وابستگی با دو اندیس و بیشتر، برابر با یک باشد- برای نمونه اگر  $\omega_{i123} = 1$  آنگاه  $\phi_{i123} = \phi_{i1}\phi_{i2}\phi_{i3}$  که استقلال سه‌تایی  $Y_{i1}, Y_{i2}, Y_{i3}$  را نتیجه می‌دهد. باید توجه داشت که اگر موفقیت و شکست برای مولفه‌هایی دو دویی معاوضه شوند (مثلا برای  $Y_{i2}$ ) و  $(1 - Y_{i2})$  مطالعه شود آنگاه ضریب همبستگی و لگ نسبت بخت در علامت تغییر می‌کنند اما در مقدار قدر مطلق تغییر نمی‌کنند اما  $\omega_{i123}$  به  $\frac{1 - \phi_{i12}\omega_{i12}}{1 - \phi_{i12}}$  تبدیل می‌شود.

علاوه بر این ضریب همبستگی و نسبت‌های بخت برای  $(1 - Y_{i2})$  و  $(1 - Y_{i1})$  برابر با  $(Y_{i2}, Y_{i1})$  است در حالی که نسبت وابستگی برای  $(1 - Y_{i2})$  و  $(1 - Y_{i1})$  تنها تابعی صعودی از نسبت وابستگی برای  $(Y_{i2}, Y_{i1})$  است.

### ۳-۴) مدل حاشیه‌ای با لگ نسبت بخت حاشیه‌ای به عنوان پارامتر همبستگی:

این مدل‌بندی توسط لیانگ و لیپسیتز و دیگران و لیپسیتز و دیگران ارائه شده است (Liang et al., 1992; Lipsitz et al., 1991). در این مدل‌بندی، در معادله ۳ پارامترهای  $\psi_i$  عباراتی از احتمال‌های حاشیه‌ای می‌باشند یعنی:

$$\psi_{ij} = \log \text{it} \{ \Pr(Y_{ij} = 1) \}$$

و پارامترهای دو طرفه  $\Omega_2$  لگ نسبت بخت حاشیه‌ای می‌باشند یعنی:

$$\exp(\omega_{rs}) = OR(Y_{ir}, Y_{is}) = \frac{\Pr(Y_{ir} = 1, Y_{is} = 1) \Pr(Y_{ir} = 0, Y_{is} = 0)}{\Pr(Y_{ir} = 0, Y_{is} = 1) \Pr(Y_{ir} = 1, Y_{is} = 0)}$$

و پارامترهای سه طرفه و بالاتر به صورت توابعی از لگ نسبت بخت حاشیه‌ای است. به عنوان

مثال:

$$\exp(\omega_{rst}) = \frac{OR(Y_{ir}, Y_{is} | Y_{it} = 1)}{OR(Y_{ir}, Y_{is} | Y_{it} = 0)}$$

یا بطور معادل:

$$\omega_{rst} = \log[OR(Y_{ir}, Y_{is} | Y_{it} = 1)] - \log[OR(Y_{ir}, Y_{is} | Y_{it} = 0)]$$

است.

با این نوع مدل‌بندی‌ها، درجه‌های مختلف وابستگی در میان  $Y_{it}$  ها قابل بیان است. به عنوان مثال مدل استقلال نتیجه می‌دهد که تمام پارامترهای همبستگی دو طرفه و بالاتر صفر است و یا اگر تمام عناصر  $\Omega_2$  مخالف صفر باشند مدل اشباع شده می‌باشد. بین این دو حالت مدل‌های مختلفی از وابستگی را می‌توانیم در نظر بگیریم. همانطور که ملاحظه می‌شود اینگونه مدل‌ها محاسبات زیادی برای برآورد پارامترها دارند (بجز برای  $n=2$ ). در نتیجه برای رهایی از این مشکل محققین اکثراً ارتباطهای مراتب بالاتر از دو را صفر در نظر می‌گیرند (Zhao & Prentice, 1990).

## ۵- مدل‌های حاشیه‌ای با استفاده از مفهوم متغیر پنهان

از مدل پروبیت چند متغیره با استفاده از مفهوم متغیر پنهان (Latent variable) برای تحلیل‌های طولی دو دویی استفاده می‌شود (Long, 1997) چون  $Y_{it}^*$  متغیری پیوسته است دسته‌ای از مدل‌های خطی را می‌توان برای تحلیل آنها در نظر گرفت، که به صورت زیر تعریف می‌شوند.

$$Y_{it}^* = x'_{it}\beta + \varepsilon_{it}$$

که در آن فرض‌هایی درباره  $\varepsilon_{it}$  ها در نظر گرفته می‌شود (Long, 1997). به عنوان مثال:

$\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im})'$  برای آزمودنی  $i$ ام، دارای توزیع نرمال  $n$  متغیره با میانگین صفر و

ماتریس کوواریانس:

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \dots & \dots & 1 \end{pmatrix} = \rho$$

در نظر گرفته می شود. این فرض مدل پروبیت چند متغیره را نتیجه می دهد. با توجه به این که  $y_{it}$  های مشاهده شده بر اساس قاعده زیر مشخص می شوند.

$$y_{it} = \begin{cases} 1 & y_{it}^* > \tau \\ 0 & y_{it}^* \leq \tau \end{cases}$$

که  $\tau = 0$  اختیار می شود (Long, 1997)، احتمال های توام و در نتیجه تابع درستنمایی را با استفاده از فرض هایی روی  $\varepsilon_i$  می توان محاسبه نمود.

همانطور که ملاحظه می شود با افزایش  $n$  تعداد پارامترهای ماتریس همبستگی افزایش چشمگیری خواهد داشت،  $\frac{n(n-1)}{2}$  پارامتر، که برآورد تمام این پارامترها نیاز به محاسبات و وقت زیاد دارد. به همین دلیل برای ماتریس همبستگی، حالت های خاصی در نظر گرفته می شود. به عنوان مثال:

$$Corr(\varepsilon_{it}, \varepsilon_{it'}) = \rho, \forall t \neq t' \quad (1)$$

یا

$$Corr(\varepsilon_{it}, \varepsilon_{it'}) = \rho^{|t-t'|} \quad (2)$$

که در حالت ۲ همبستگی بین دو پاسخ آزمودنی یکسان با افزایش فاصله زمانی بین دو پاسخ کاهش می یابد.

## ۶- رهیافت معادلات برآورد تعمیم یافته

این رهیافت توسط لیانگ و زیگرپیشنهاد شده است (Liang and Zeger, 1981). رهیافت معادلات برآورد تعمیم یافته، رهیافت واحدی را برای تحلیل برآمدهای گوناگون پیوسته و گسسته پیشنهاد می کند. در این میان دسته ای از معادلات برآوردگر تعمیم یافته (GEE) برای پارامترهای رگرسیون پیشنهاد شده است. در این رهیافت تنها فرض هایی که اختیار می شود فرض درباره امید حاشیه ای مرتبه اول و دوم پاسخ ها است و درباره توزیع کامل آنها فرضی اختیار نمی شود.



این رهیافت برای پارامترهای رگرسیون برآوردهای سازگار می دهد. دراصل (GEE) حالت چند متغیره شبه درستنمایی (Quasi likelihood) است که توسط ودربرن (Wedderburn, 1974) بیان شده است.

دراین رهیافت تابعی از امید حاشیه ای متغیر وابسته به صورت تابعی خطی از متغیرهای کمکی مشخص می شود. همچنین فرض می شود که واریانس تابعی معلوم از میانگین حاشیه ای است. به علاوه ماتریس همبستگی عملی برای مشاهدات هر آزمودنی، معین می شود. یعنی:

$$var(Y_{ij}) = v(\mu_{ij}) / \phi \quad \text{و} \quad h(\mu_{ij}) = x'_{ij} \beta$$

و همبستگی عملی مشاهدات برای آزمودنی  $i$  ام  $R_i(\alpha)$  فرض می شود که به بردار  $s \times 1$  از پارامترهای اضافی،  $\alpha$ ، وابسته است.  $\alpha$ ، همبستگی پاسخ های آزمودنی  $i$  ام را به طور کامل بیان می کند. اکثراً از  $R(\alpha)$  به عنوان ماتریس همبستگی عملی استفاده می شود، زیرا انتظار این که ماتریس همبستگی واقعی به درستی مشخص شود نیست. در نتیجه ماتریس کوواریانس عملی  $Y_i$  به صورت:

$$V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2} / \phi$$

است که  $A_i$  ماتریس قطری  $n_i \times n_i$  با عناصر  $(\mu_{ij})$  در  $j$  امین عنصر قطری آن می باشد. معادلات برآورد به صورت زیر می باشد.

$$\sum_{i=1}^m D_i' V_i^{-1} S_i = 0 \quad (4)$$

که در آن  $S_i = Y_i - \mu_i$  و  $D_i = \frac{\partial \mu_i}{\partial \beta}$  است. هنگامی که  $m, \dots, i, \dots, 1, n_i = 1$  معادله (4)

به معادله شبه درستنمایی تحویل می شود (Wedderburn, 1974).

معادله ۴ می تواند به صورت تنها تابعی از  $\beta$  بیان شود به این صورت که برآوردهای سازگار  $\alpha$  و  $\phi$  در معادله ۴ قرار گیرد (لیانگ و زیگر، ۱۹۸۶) یعنی به صورت زیر خواهد بود.

$$\sum_{i=1}^m U_i [\beta, \hat{\alpha} \{ \beta, \hat{\phi}(\beta) \}] = 0 \quad (5)$$

$\hat{\beta}_i$  که نشان دهنده برآورد  $\beta$  تحت رهیافت (GEE) است، جواب معادله بالا خواهد بود. براساس قضیه ۲ در لیانگ و زیگر (۱۹۸۶)،  $m^{1/2}(\hat{\beta}_i - \beta)$  بطور مجانبی گاوسی با میانگین صفر و ماتریس کوواریانس  $V_i$  به صورت زیر است:

$$\begin{aligned}
 V_G &= \lim_{m \rightarrow \infty} m \left( \sum_{i=1}^m D_i V_i^{-1} D_i \right)^{-1} \\
 &\times \left\{ \sum_{i=1}^m D_i V_i^{-1} \text{cov}(Y_i) V_i^{-1} D_i \right\} \left( \sum_{i=1}^m D_i V_i^{-1} D_i \right)^{-1} \\
 &= \lim_{m \rightarrow \infty} m V_1^{-1} V_1^{-1}
 \end{aligned}$$

برآورد واریانس  $\hat{V}_G$  -  $\hat{\beta}_G$  می‌تواند برای یافتن واریانس برآوردگرها استفاده شود (Zeger & Liang, 1992).

### ۷- مقایسه کلی مدل‌های حاشیه‌ای

در این بخش به بررسی و مقایسه مدل‌های حاشیه‌ای ارائه شده در بخش‌های قبل می‌پردازیم. ابتدا مدل استقلال، که فرض بر عدم همبستگی بین پاسخ‌های متعلق یک آزمودنی خاص دارد، را در نظر بگیرید. براساس قضیه ۱ لیانگ و زیگر (۱۹۸۶)،  $\hat{\beta}_i$  (برآورد پارامتر تحت مدل استقلال) برآوردی سازگار برای  $\beta$  مهیا می‌کند. اما این روش استنباط‌هایی نادرست درباره پارامترها می‌دهد زیرا فیتزموریس و دیگران (۱۹۹۳) بیان کردند که اکثراً خطای استاندارد برآورد پارامترهای مربوط به متغیرهای زمان مانا میل به کم برآورد شدن و برای پارامترهای مربوط به متغیرهای زمان متغیر میل به بیش برآورد شدن دارند. و بنابراین برآوردهای واریانس‌ها سازگار نیستند.

رهیافت‌هایی که در بخش چهارم بیان شدند هرکدام دارای محاسن و معایبی می‌باشند. پارامترهای همبستگی به صورت توابعی از لگ نسبت بخت شرطی بر خلاف همبستگی‌های بهادر و لگ نسبت بخت حاشیه‌ای مقید به میانگین حاشیه‌ای نیستند. ولی این نوع پارامترها به  $n$  (تعداد پاسخ‌های تکرار شده برای هر آزمودنی) وابسته می‌باشند.

وقتی پارامترهای همبستگی به صورت توابعی از لگ نسبت بخت حاشیه‌ای بیان شود، پارامترهای همبستگی به  $n$  وابسته نیست. و به همین دلیل برای تعداد پاسخ‌های تکرار شده متفاوت برای آزمودنی‌های مختلف، این مدل کاربرد دارد. دیگر مزیت مدل لگ نسبت بخت حاشیه‌ای نسبت به مدل با لگ نسبت بخت شرطی آن است که قابل دوباره تولید شدن می‌باشد ولی مدل با لگ نسبت بخت شرطی قابلیت دوباره تولید شدن را ندارد (مدلی قابل دوباره تولید شدن است که اگر  $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{im})'$  در یک مدل حاشیه‌ای صدق کند، آنگاه زیر مجموعه‌ای

از  $Y_i$  مانند  $Y_i' = (Y_{i1}, \dots, Y_{im})'$  نیز در مدل با پارامترهای متناظر آن صدق کند).

هنگامی که پارامترهای همبستگی بین پاسخ‌ها براساس نسبت وابستگی بیان می‌شود، پارامترهای همبستگی همانند لگ نسبت بخت حاشیه‌ای وابسته به  $\eta$  نیست. باید توجه داشت که پیچیدگی قید پارامترهای همبستگی نسبت‌های وابستگی و لگ نسبت بخت حاشیه‌ای به میانگین‌ها کمتر از پیچیدگی قید همبستگی‌های بهادر است.

در هر سه مدل بخش ۴، ضرایب رگرسیونی و پارامترهای همبستگی بر هم عمود می‌باشند. این بدین معنی است که کنواریانس برآوردگرهای ضرایب رگرسیونی و پارامترهای همبستگی برابر صفر است. (بارندورف نیلسون و کاکس، ۱۹۹۴).

برآوردهای ماکسیمم درست‌نمایی پارامترهای مدل‌های ذکر شده در بخش چهارم براساس روشهای تکراری عددی بدست می‌آیند. برای تمام این مدل‌ها بدون توجه به ساختار همبستگی، برآورد پارامترهای رگرسیونی سازگار می‌باشند ولی تنها اگر ساختار همبستگی برای این مدل‌ها بدرستی مشخص شده باشند آنگاه برآورد  $\text{cov}(\hat{\beta})$  نیز برآوردی سازگار می‌باشد.

رهیافت GEE که رهیافتی غیردرست‌نمایی است برخلاف مدل پروبیت چند متغیره که به فرض توزیع چند متغیره نرمال برای متغیرهای پنهان نیاز دارد، تنها به فرضیهایی درباره دو گشتاور اول پاسخ‌ها نیاز دارد. البته باید توجه کرد که در توزیع گاوسی، فرض درباره دو گشتاور اول، توزیع را به طور کامل مشخص می‌کند درحالی که برای داده‌های دودویی که مورد بحث اصلی است، این مطلب برقرار نمی‌باشد (توجه کنید که متغیرهای پنهان ما دارای توزیع گاوسی‌اند نه متغیرهای مشاهده شده). در GEE تابع پیوندهای مختلفی همچون پروبیت، لجیت و ... می‌تواند مورد استفاده قرار گیرد، ولی در پروبیت چند متغیره تنها پیوند پروبیت و در مدل برای لگ نسبت بخت حاشیه‌ای تنها پیوند لجیت مورد استفاده است. GEE همانند مدل با نسبت وابستگی و نسبت بخت حاشیه‌ای، برای داده‌های با اندازه تکرارهای متفاوت برای آزمودنی‌های مختلف قابل اجرا است.

یکی از محدودیتهای GEE نداشتن روش‌هایی برای آزمون نیکویی برازش، مقایسه مدل و استنباط بر روی پارامترها بر اساس درست‌نمایی است زیرا در این روش، توزیع توام به طور کامل مشخص نمی‌شود. با توجه به این که برآوردگرها دارای توزیع نرمال مجانبی هستند، در این روش برای حل این مشکل از آماره‌های والد (Wald Statistics) استفاده می‌شود. یکی دیگر از محدودیتهای GEE آن است که اگر مدل برای میانگین‌های حاشیه‌ای به درستی مشخص نشده باشد، برآوردهای ناسازگار برای  $\beta$  به دست می‌آید.

### ۸- بررسی مانده‌ها

بررسی مانده‌ها برای نیکویی برازش مدل و یافتن دورافتاده‌ها (مشاهداتی که روی نیکویی برازش مدل تاثیر می‌گذارند) در حالت مطالعات طولی، همانند مطالعات مقطعی ضروری است. تفاوتی که در این حالت برای مانده‌ها نسبت به مطالعات مقطعی وجود دارد این است که برای هر آزمودنی چند پاسخ داریم. برای محاسبه مانده‌ها اگر  $\Sigma_i(\beta) = \text{diag}\{\text{var}(Y_{ij})\}$  در نظر گرفته شود آنگاه همبستگی بین پاسخ‌های متعلق به یک آزمودنی خاص در نظر گرفته نمی‌شود. اما اگر از ماتریس کوواریانس  $Y_i$  استفاده شود، این همبستگی برای محاسبه مانده‌ها نیز مورد استفاده قرار می‌گیرد. در این حالت مانده‌های پیروان برای مشاهدات  $Y_i$  را به صورت زیر معرفی می‌کنیم. مانده‌ها به صورت:

$$r_i^p = \sum_i^{-1/2}(\hat{\beta})(y_i - \hat{\mu}_i)$$

داده می‌شود که در آن  $\sum_i(\hat{\beta})$  برآورد ماتریس کوواریانس مشاهدات است. همچنین آماره نیکویی برازش پیروان به صورت:

$$\chi^2 = \sum_{i=1}^m \chi_p^2(y_i, \hat{\mu}_i)$$

است که در آن:

$$\chi_p^2(y_i, \hat{\mu}_i) = (y_i - \hat{\mu}_i)' \Sigma_i^{-1}(\hat{\beta})(y_i - \hat{\mu}_i)$$

$\chi_p^2(y_i, \hat{\mu}_i)$  را می‌توان به صورت:  $r_i^p{}' r_i^p$  که در آن  $r_i^p$  ترانزاده بردار  $r_i^p$  تعریف شده در بالا است، نوشت. بزرگ بودن  $\chi_p^2$  در مقایسه با درجه آزادی آن (تعداد پارامترهای برآورد شده -  $m$ ) گواه بر بدی برازش مدل دارد. در این صورت یا مدل مورد استفاده مناسب نیست یا، در صورت وجود، مؤلفه‌های دیگری بایستی در مدل به‌عنوان متغیرهای تبیینی در نظر گرفته شود. به‌عنوان مثال با مدل دو متغیره پروبیت با دو پاسخ برای هر آزمودنی به جای ساختار

کوواریانس زیر که فرض استقلال را در نظر می‌گیرد:

$$\Sigma(\beta) = \begin{bmatrix} \pi_{11}(1-\pi_{11}) & 0 \\ 0 & \pi_{12}(1-\pi_{12}) \end{bmatrix}$$

ماتریس کوواریانس  $Y_i$  به صورت زیر را در نظر می‌گیریم:

$$\Sigma_i(\beta) = \begin{bmatrix} \pi_{11}(1-\pi_{11}) & \pi_{112} - \pi_{11}\pi_{12} \\ \pi_{112} - \pi_{11}\pi_{12} & \pi_{12}(1-\pi_{12}) \end{bmatrix}$$

که در آن:

$$\text{cov}(Y_{i1}, Y_{i2}) = \pi_{i12} - \pi_{i1}\pi_{i2}$$

با:

$$\begin{aligned} \pi_{i12} &= \Pr(Y_{i1} = 1, Y_{i2} = 1) \\ &= 1 - \Phi(-x'_{i1}\beta_1) - \Phi(-x'_{i2}\beta_2) + \Phi_2(-x'_{i1}\beta_1, -x'_{i2}\beta_2, \rho) \end{aligned}$$

و

$$\pi_{ij} = \Pr(Y_{ij} = 1) = \Phi(x'_{ij}\beta) \quad j=1,2$$

$\Phi(\cdot)$  تابع توزیع نرمال دو متغیره با همبستگی  $\rho$  بین دو متغیر است و  $\Phi_2(\cdot, \dots, \rho)$  تابع توزیع نرمال استاندارد است. اگر برآورد  $\beta$  را در  $\sum_i^{-1/2}(\beta)$  قرار دهیم برآورد این ماتریس بدست می‌آید. همان طور که ملاحظه می‌شود برآورد پارامتر  $\rho$  در محاسبه برآورد ماتریس کوواریانس و بنابراین مانده‌ها نقش دارد.

#### ۹- مثال عملی: (برازش مدل‌های حاشیه‌ای بر روی داده‌های آسم)

این داده‌ها (رتنیتزکی و وایپیچ، ۱۹۹۴) در ۶ شهرستان منطقه هاروارد، ماساچوست جمع‌آوری شده‌اند، به این ترتیب که ۷۰۶ پسر و ۷۱۳ دختر سفید پوست را در ۹ سالگی و دوباره در ۱۳ سالگی از نظر ابتلا یا عدم ابتلا به آسم مورد مطالعه قرار داده‌اند. می‌خواهیم تاثیر زمان و جنس را بر احتمال داشتن آسم مورد آزمون قرار دهیم. جدول ۱ داده‌های کامل و بدون مقادیر گمشده برای متغیر پاسخ در این داده‌ها را نشان می‌دهد. در بررسی‌های جدا برای رد آزمون کملاً تصادفی بودن مقادیر گمشده گواهی بدست نیامد (Little and Rubin, 1987). همانطور که جدول ۱ نشان می‌دهد ۵۵۷ نفر پسر و ۵۹۰ نفر دختر به متغیر مورد علاقه پاسخ داده‌اند. برای بررسی تاثیر سن و زمان بر احتمال داشتن آسم، مدل‌های حاشیه‌ای که قبلاً ذکر شد را برازش می‌دهیم.

مدل برای نسبت وابستگی، لگ نسبت بخت حاشیه‌ای و مدل استقلال به صورت زیر در نظر گرفته شده است:

$$\text{logit}(\pi_{ij}) = a_j + b_j \text{sex}_i, \quad j=1,2, \quad i=1, \dots, 1147$$

که در آن  $\pi_{ij}$  احتمال داشتن آسم فرد  $i$  ام در زمان  $j$  ام است،  $\text{sex}_i$  جنس آزمودنی  $i$  ام (۱ : پسر، ۰ : دختر) را نشان می‌دهد و  $\text{var}(Y_{ij}) = \pi_{ij}(1 - \pi_{ij})$  است. پارامتر همبستگی برای نسبت وابستگی با لگ نسبت وابستگی و با لگ نسبت بخت حاشیه‌ای برای

مدل نسبت بخت حاشیه‌ای بیان می‌شود. مدل حاشیه‌ای پروبیت دو متغیره با مفهوم متغیر پنهان نیز به این داده‌ها برازش داده‌ایم. مدل در نظر گرفته شده به صورت:

$$Y_{ij}^* = a_j + b_j sex_j + \varepsilon_{ij} \quad j=1,2$$

است. که  $Y_{i1}^*$  و  $Y_{i2}^*$  متغیرهای پنهان هستند. همچنین همانطور که در بخش ۵ ذکر شد  $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2})'$  برداری با توزیع نرمال دو متغیره است.

جدول ۱- داده‌های مربوط به آسم

		۱۳سالگی			
		آسم دارد	آسم ندارد	جمع کل	
۹ سالگی	پسرها	آسم دارد	۲۲	۶	۲۸
		آسم ندارد	۱۵	۵۱۴	۵۲۹
		جمع کل	۳۷	۵۲۰	۵۵۷
	دخترها	آسم دارد	۱۳	۳	۱۶
		آسم ندارد	۱۳	۵۶۱	۵۷۴
		جمع کل	۲۶	۵۶۴	۵۹۰

نتایج برازش مدل‌ها در جدول ۲ آمده است. با استفاده از نتایج بدست آمده در جدول ۲ مشاهده می‌شود که پارامترهای همبستگی وجود همبستگی بالای مثبت ( $\rho$  - مقدار = ۰/۰۰۰) را بین دو پاسخ برای هر آزمودنی نشان می‌دهد.

همان‌گونه که ملاحظه می‌شود تمام این مدل‌ها ضریب ثابت را معنی‌دار نشان می‌دهند یعنی عواملی مانند زمان بوده‌اند که در نظر نگرفته‌ایم. به این دلیل دوباره مدل GEE را بر روی داده‌ها برازش داده‌ایم. مدل GEE به صورت زیر تعیین می‌گردد:

$$\logit(\pi_{ij}) = a_1 + b_1 sex_j + cI(age13)_{ij} \quad j=1,2$$

که در آن  $I(age13)$  به صورت زیر تعریف می‌شود:

$$I(age13) = \begin{cases} 1 & \text{اگر فرد سیزده ساله باشد،} \\ 0 & \text{اگر فرد سیزده ساله نباشد،} \end{cases}$$

در GEE ماتریس همبستگی به صورت غیر ساختاری در نظر گرفته شده است. نتایج برآزش مدل‌ها در جدول ۲ آمده است.

هرچند مقایسه مدل‌ها در یک مثال امکان پذیر نیست و نیاز به بررسی‌های نظری الزامی است، خلاصه‌ای از عملکرد مدل‌های مختلف در برآورد پارامترها می‌آوریم. در همه مدل‌ها گواه کافی بر معنی‌داری پارامتر منتسب به زمان ( $c$  یا تفاضل  $a_1, a_2$ ) وجود دارد. به این معنی که هر چه زمان اقامت در این ۶ شهرستان بیشتر می‌شود احتمال ابتلا به آسم افزایش می‌یابد. در مدل حاشیه‌ای پروبیت دو متغیره جنس در زمان اول معنی‌دار است ولی در زمان دوم معنی‌دار نیست. به این معنی که پسران در سنین پایین برای آسم داشتن محتمل‌تر از دختران هستند. ولی با گذشت زمان پسران بهبود یافته بیشتر از دخترانی هستند که به آسم مبتلا می‌شوند. مدل پروبیت با فرض استقلال نیز به نتیجه یکسانی می‌رسد، اما واریانس‌های برآوردگرهای پارامترها کمی بیش یا کم برآورد می‌شوند. مدل استقلال با پیوند لجیت نیز در مورد برآورد پارامترها به نتیجه‌ای شبیه به مدل پروبیت دو متغیره می‌رسد، ولی این مدل در مقایسه با مدل نسبت وابستگی با پیوند لجیت یا نسبت بخت با پیوند لجیت واریانس برآوردگرها را بسیار متزلزل برآورد می‌کند.

جدول ۲- برآورد پارامترها تحت مدل‌های در نظر گرفته شده.

پارامتر	نسبت بخت با پیوند لجیت		نسبت وابستگی با پیوند لجیت		مدل استقلال با پیوند لجیت	
	برآورد	خطای استاندارد	برآورد	خطای استاندارد	برآورد	خطای استاندارد
$a_1$	-۵۱۰۶۰	۰/۴۱۸	-۳/۳۹۰	۰/۲۳۲	-۳/۵۸۰	۰/۲۵۲
$a_2$	-۳/۷۲۳	۰/۲۶۰	-۲/۸۸۸	۰/۱۷۱	-۳/۰۷۷	۰/۲۰۰
$b_1$	۰/۵۲۴	۰/۴۱۳	۰/۲۷۸	۰/۲۶۲	۰/۶۴۱	۰/۳۱۹
$b_2$	۰/۱۵۳	۰/۳۴۲	۰/۰۷۳	۰/۱۸۶	۰/۴۲۴	۰/۲۶۳
پارامتر همبستگی	۴/۹۸۶	۰/۴۲۱	۱۴/۲۴۱	۱/۶۴۵	----	----
منهای لگاریتم درستمایی	۳۳۷/۲۱۱		۳۳۸/۷۴۶		۴۲۷/۱۸۱	
پارامتر	مدل GEE پیوند لجیت		مدل پروبیت		مدل استقلال با پیوند پروبیت	
	برآورد	خطای استاندارد	برآورد	خطای استاندارد	برآورد	خطای استاندارد
$a_1$	-۳/۴۸۶	۰/۲۱۱	-۱/۹۳۵	۰/۱۰۵	-۱/۹۳۷	۰/۱۰۶
$a_2$	----	----	-۱/۷۰۵	۰/۰۹۰	-۱/۷۰۵	۰/۰۹۱
$b_1$	۰/۴۸۹	۰/۲۵۵	۰/۲۸۳	۰/۱۳۵	۰/۲۸۳	۰/۱۳۹
$b_2$	----	----	۰/۲۰۲	۰/۱۲۱	۰/۲۰۲	۰/۱۲۲
$c$	۰/۳۷۷	۰/۱۲۱	----	----	----	----
پارامتر همبستگی	۰/۶۴۴	----	۰/۹۲۶	۰/۲۴۵	----	----
منهای لگاریتم درستمایی	-----		۳۳۷/۱۰۶		۴۲۷/۱۸۱	

در مدل‌های نسبت وابستگی با پیوند لوجیت یا نسبت بخت با پیوند لوجیت اثر جنس در هیچیک از زمانها معنی‌دار نیست، به این معنی که پسرها و دخترها در هر سنی به یک‌اندازه برای داشتن آسم محتملند. در مدل GEE، که در آن بر خلاف مدل‌های دیگر هیچ فرضی در مورد توزیع پاسخها اعمال نمی‌شود، گواهی قوی بر معنی‌داری اثر زمان ولی گواهی ضعیف بر معنی‌داری جنس وجود دارد.

• برای نیکویی برازش مدل بر اساس مانده‌ها که در بخش ۸ بیان شده است، می‌توانیم عمل کنیم. مثلاً برای مدل پروبیت دو متغیره برازش شده بر روی این داده‌ها، آماره کی دو برابر با  $2543/089$  است که با این مقدار آماره فرض مناسب بودن مدل رد می‌شود. اگر در مدل پروبیت دو متغیره همبستگی را در نظر نگیریم مقدار آماره برابر با  $2558/084$  است و خوبی برازش این مدل نیز رد می‌شود. فکر می‌کنیم که عدم برازش خوب به دلیل وجود عامل‌های متعدد موثر دیگری همچون میزان آلودگی هوا و عوامل محیطی و ژنتیکی دیگر است، که در این بررسی در نظر گرفته نشده است. بهر حال کاهش مقدار آماره کی دو در مدل عدم استقلال نسبت به مدل استقلال به خاطر در نظر گرفتن همبستگی بین پاسخها است.

## Reference

- Agresti, A. (2002) *Categorical Data Analysis*. New York: John Wiley.
- Bahadur, R.R. (1961) *A representation of the joint distribution of responses to  $n$  dichotomous item. In studies on item analysis and prediction*. Stanford: Stanford University Press.
- Barndorff-Nielson, O.E., and Cox, D.R. (1994) *Inference and Asymptotics*. London: Chapman & Hall.
- Diggel, P.J., Liang, K., Zeger, S. (1994) *Analysis of Longitudinal Data*. Oxford Science Publication.
- Ekhholm, A., Smith, P.W.F., McDonald, J.W. (1995) Marginal regression analysis of a multivariate binary response. *Biometrika*, **82**, 847-854.
- Fitzmaurice, G.M., Laird, N.M. (1993) *A Likelihood-Based method for analysing longitudinal binary responses*. *Biometrika*, **80**, 141-151.
- Fitzmaurice, G.M., Laird, N.M. Rotnitzky, A.G. (1993) *Regression Models for Discrete Longitudinal responses*. *Statistical Science*, **8**, 284-309.



- Liang, K.Y., Zeger, S.L. (1986) *Longitudinal data analysis using generalized linear models*. Biometrika, **73**, 13-22.
- Liang, K.Y., Zeger, S.L., Qaqish, B. (1992) *Multivariate regression analysis for categorical data ( with discussion )*. Journal of the Royal Statistical Society, B, **54**, 3-40.
- Lipsitz, S., Laird, N., Harrington, D. (1991) *Generalized estimating equations for correlated binary data: Using odds ratios as a measure of association*. Biometrika, **78**, 153-160.
- Little, R. J. and Rubin, D. (1987) *Statistical analysis with missing data*. New York: Wiley. .
- Long, S.J. (1997) *Regression models for categorical and limited dependent variables*. London: SAGE.
- Rotnitzky, A. and Wypij, D. (1994) *A note on the Bias of Estimation with Missing Data*. Biometrika, **147**, 87-99.
- Wedderburn, R.W.M. (1974) *Quasi-likelihood functions, Generalized linear Models, and the Gauss-Newton method*. Biometrika, **61**, 439-447.
- Zeger, S.L., Liang, K.Y. (1992) *An overview of methods for the analysis of longitudinal data*. Statistics in medicine, **11**, 1825-1839.
- Zhao, L.P., Prentice, R.L. (1990) *Correlated binary regression using a quadratic exponential model*. Biometrika, **77**, 642-648.