فصل ۶

مطالعات طولی: داده‌های دارای پاسخ دودویی

در این فصل به معرفی و بررسی مطالعات طولی با پاسخ دودویی می‌پردازیم. برای این منظور، در بخش اول، مفهومی از مطالعات طولی برای پاسخ‌های دودویی، همراه با نمادگذاری در این داده‌ها بیان می‌شود. در بخش دوم، توزیع‌های مورد استفاده برای این نوع داده‌ها معرفی می‌شوند. مدل‌های حاشیه‌ای، انتقالی و اثرهای تصادفی برای داده‌های دارای پاسخ دودویی طولی را در بخش سوم ارائه می‌کنیم. در بخش چهارم، روش‌های محاسباتی برای براورد پارامترهای مدل‌های معرفی‌شده ارائه می‌شوند. مثال‌های کاربردی با پاسخ دودویی در بخش پنجم خواهند آمد.

۱.۶ مفاهیم و نمادگذاری

از مشخصه‌های بارز مطالعات طولی، اندازه‌گیری‌های مکرر برای آزمودنی‌های مختلف در طول زمان است. مطالعات طولی در مقابل مطالعات مقطعی‌اند، چرا که در مطالعات مقطعی، برای هر آزمودنی، فقط یک پاسخ اندازه‌گیری می‌شود. اغلب ممکن است به

یک سؤال علمی برخورد کنیم که هم با مطالعات طولی قابل پاسخ‌گویی است و هم با مطالعات مقطعی؛ با این تفاوت که در مطالعات طولی، اثرهای سن و اثرهای گروه، مورد بررسی قرار می‌گیرند. نکته‌ی قابل توجه این است که در مطالعات طولی می‌توان تغییرات و اثر زمان (اثر سن) را برای هر آزمودنی، از اثر گروه که تفاوت‌های آزمودنی‌ها در سطح پایه است (اثرهای گروهی) جدا کرد؛ در حالی که بر خلاف مطالعات طولی، در مطالعات مقطعی قابلیت بیان این اختلاف‌ها وجود ندارد. داده‌های طولی به‌صورت آینده‌نگر و مطابق موضوعات رو به جلو در زمان، یا به‌صورت گذشته‌نگر به‌وسیله‌ی استخراج چندین اندازه برای هر آزمودنی در گذشته، مانند گزارش‌های بیمارستانی دسته‌بندی می‌شوند. از آن‌جا که داده‌های گذشته نامرغوب‌اند، داده‌های طولی اغلب آینده‌نگر هستند.

در مطالعات طولی، همبستگی بین پاسخ‌ها در طول زمان، مورد بررسی قرار می‌گیرد. تفاوت مطالعات طولی با سری‌های زمانی که آن نیز نوعی همبستگی بین پاسخ‌ها در طول زمان را مورد بررسی قرار می‌دهد، این است که در مطالعات طولی، برای هر آزمودنی (که تعداد آزمودنی‌ها زیاد است) در طول زمان، چندین پاسخ (که تعداد آن‌ها چندان زیاد نیست) اندازه‌گیری می‌شود، در حالی که در سری‌های زمانی فقط پاسخ برای یک آزمودنی (اما در تعداد نسبتاً زیادی) در نظر گرفته می‌شود. در این فصل کتاب، داده‌های طولی با پاسخ دودویی مورد بحث و بررسی قرار می‌گیرند.

به‌منظور آشنایی با مطالعات طولی که از این به بعد با آن‌ها روبه‌رو خواهیم بود، اکنون به معرفی نمادگذاری مطالعات طولی می‌پردازیم. فرض کنید Y_{it} متغیر پاسخ و x_{it} برداری p—بعدی از متغیرهای توضیحی مشاهده‌شده در زمان t برای $t = ۱, \ldots, T_i$ و $i = ۱, \ldots, n$ است که در آن i بیانگر آزمودنی است و t، زمان را نشان می‌دهد.

به‌عنوان مثال، اگر فرض کنیم که تعداد مشاهدات برای همه‌ی آزمودنی‌ها یکسان است ($T_i = T$ به‌ازای همه‌ی iها)، برامدهای پاسخ به‌صورت زیر نمایش داده می‌شوند:

	$t = ۱$	$t = ۲$	$\cdots$	$t = T$
$i = ۱$	$Y_{۱۱}$	$Y_{۱۲}$	$\cdots$	$Y_{۱T}$
$i = ۲$	$Y_{۲۱}$	$Y_{۲۲}$	$\cdots$	$Y_{۲T}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i = n$	$Y_{n۱}$	$Y_{n۲}$	$\cdots$	Y_{nT}

میانگین و واریانس Y_{it} را به‌ترتیب با $\mu_{it} = E(Y_{it})$ و $v_{it} = \text{var}(Y_{it})$ نشان می‌دهیم. مجموعه‌ی برامدهای تکرار شده برای آزمودنی iام را با یک بردار به طول T_i به‌صورت زیر نشان می‌دهیم:

$$Y_i = (Y_{i۱}, Y_{i۲}, \ldots, Y_{iT_i})'.$$

میانگین این بردار را با $\mu_i = E(Y_i)$ و ماتریس واریانس- کوواریانس آن را با $V_i = \text{var}(Y_i)$ نشان می‌دهیم که دارای مرتبه‌ی $T_i \times T_i$ است. درایه‌ی (j,k) در V_i، کوواریانس بین Y_{ij} و Y_{ik} را نشان می‌دهد، که آن را با v_{ijk} نمایش خواهیم داد. همچنین از R_i به‌عنوان ماتریس همبستگی Y_i استفاده خواهیم کرد.

۲.۶ توزیع مناسب برای داده‌های طولی دارای پاسخ دودویی

همان‌طور که در فصل ۲ به‌تفصیل به آن پرداختیم، توزیع مورد استفاده برای پاسخ‌های دودویی، برنولی است. حال با توجه به این‌که نوع مطالعات طولی است، مسلماً تأثیری در توزیع‌های حاشیه‌ای مورد استفاده برای پاسخ‌های دودویی به وجود نمی‌آورد (هرچند توزیع بردار Y_i چندمتغیره است) و فقط مدل‌های مورد استفاده برای تحلیل داده‌ها از دیدگاه توزیع مشترک متفاوت‌اند که در زیر به‌تفصیل به آن می‌پردازیم.

۳.۶ مدل‌های مختلف برای تحلیل داده‌های طولی دارای پاسخ دودویی

در این بخش به بررسی مدل‌های مورد استفاده برای داده‌های طولی با پاسخ‌های دودویی بدون مقادیر گم‌شده می‌پردازیم. در زیربخش اول، رهیافت‌های معمول برای تحلیل داده‌های طولی را بیان می‌کنیم. در این زیربخش، نخست مدل حاشیه‌ای را ارائه می‌کنیم. برای بررسی همبستگی بین پاسخ‌ها در این مدل از سه روش لگاریتم نسبت بخت

حاشیه‌ای، لگاریتم نسبت بخت شرطی و لگاریتم نسبت بخت وابستگی استفاده می‌شود. سپس در زیربخش‌های بعد، رهیافت‌های دیگر در تحلیل داده‌های طولی بیان می‌شود، که شامل مدل‌های اثرهای تصادفی و انتقالی است.

۱.۳.۶ مدل‌های حاشیه‌ای

مدل‌های حاشیه‌ای، مدل‌هایی هستند که در آن‌ها برای یک آزمودنی معین، تأثیر متغیرهای تبیینی روی پاسخ‌ها به‌طور مجزا از همبستگی بین پاسخ‌ها، مدل‌بندی می‌شود. در این مدل‌ها امید ریاضی حاشیه‌ای $E(Y_t|\boldsymbol{x}_t)$ به‌صورت تابعی از متغیرهای تبیینی مدل‌بندی می‌شود. با فرض $T_i = T$ مشخصات مدل‌های حاشیه‌ای به‌صورت زیر است:

آ) میانگین حاشیه‌ای یا احتمال پاسخ که به‌صورت زیر تعریف می‌شود، به‌وسیله‌ی مدل‌های دارای پاسخ دودویی مشخص می‌شود.

$$\Pr(Y_{it} = ۱|\boldsymbol{x}_{it}) = E(Y_{it}|\boldsymbol{x}_{it}), \qquad t = ۱, \ldots, T$$

و

$$\pi_{it}(\beta_{it}) = E(Y_{it}|x_{it}) = g_{it}(z'_{it}\beta_{it})$$

است که در آن g_{it} یک تابع پیوند مانند لوجیت است و $z_{it} = z_{it}(x_{it})$ یک بردار طرح است.

ب) واریانس حاشیه‌ای که به میانگین حاشیه‌ای وابسته است، به‌وسیله‌ی تابع واریانس زیر مشخص می‌شود:

$$\mathrm{var}(Y_{it}|\boldsymbol{x}_{it}) = \epsilon(\pi_{it}) = \pi_{it}(۱ - \pi_{it}).$$

پ) کوواریانس بین Y_j و Y_k که تابعی از میانگین‌های حاشیه‌ای و پارامتر اضافی α با تابع معلوم c است، به‌صورت زیر مشخص می‌شود:

$$\mathrm{cov}(Y_{ij}, Y_{ik}) = c(\pi_{ij}, \pi_{ik}; \alpha).$$

به‌منظور بررسی همبستگی بین پاسخ‌ها برای آزمودنی iام، تابع چگالی مشترک پاسخ‌ها به‌صورت زیر بیان می‌شود:

$$f(y_i; \Psi_i, \Omega) = \exp\{\Psi'_i y_i + \Omega'_i w_i - A(\Psi_i, \Omega_i)\} \qquad (۱.۶)$$

که در آن $y_i = (y_{i1}, \dots, y_{iT})'$ و $w_i = (y_{i1}y_{i2}, \dots, y_{iT-1}y_{iT}, \dots, y_{i1}y_{i2}\dots y_{iT})'$ برداری از حاصل‌ضرب‌های دوتایی و بیش‌تر است، و

$$\Psi_i = (\psi_{i1}, \dots, \psi_{iT})'$$

و

$$\Omega_i = (w_{i12}, \dots, w_{i(T-1)T}, \dots, w_{i12\dots T})'$$

بردارهایی از پارامترهای متعارف هستند.

اکنون به معرفی سه روش برای بررسی همبستگی بین پاسخ‌ها می‌پردازیم که در آن فرض کرده‌ایم برای هر i داریم $T_i = T$. شایان ذکر است که این همبستگی‌ها همراه با معادلات درست‌نمایی در زیربخش ۳.۴.۶ به‌تفصیل معرفی خواهند شد.

لگاریتم نسبت بخت شرطی

در این مدل‌بندی، در معادله‌ی (۱.۶) پارامترهای Ψ_i عبارت‌هایی از احتمال‌های شرطی به‌صورت $\psi_{it} = \text{logit}\{\Pr(Y_{it} = 1 | Y_{is} = 0, s \neq t)\}$ هستند و پارامترهای Ω_i عبارت‌هایی از لگاریتم نسبت بخت شرطی‌اند؛ یعنی

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

لگاریتم نسبت بخت وابستگی

در این مدل‌بندی، در معادله‌ی (۱.۶) ارتباط و همبستگی با نسبت وابسته مدل‌بندی می‌شود، که عبارت‌هایی از پارامتر میانگین ϕ می‌باشند. اگر به‌عنوان مثال، پاسخ دودویی دومتغیره برای آزمودنی iام، (Y_{i1}, Y_{i2})، با پارامتر میانگین $\phi_i = (\phi_{i1}, \phi_{i2}, \phi_{i12})$ در نظر گرفته شود، نسبت وابسته، w_{i12}، و لگاریتم نسبت وابسته، λ_{i12}، به‌صورت زیر تعریف

می‌شوند:

$$w_{i12} = \frac{\phi_{i12}}{\phi_{i1}\phi_{i2}}, \qquad \lambda_{i12} = \log(w_{i12})$$

که در آن

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

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

و

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

به همین ترتیب، نسبت‌های وابسته‌ی مراتب بالاتر نیز به‌صورت زیر تعریف می‌شوند:

$$w_{i123} = \frac{\phi_{i123}}{\phi_{i1}\phi_{i2}\phi_{i3}}, \ldots, w_{i123\cdots n} = \frac{\phi_{i123\ldots T}}{\phi_{i1}\phi_{i2}\cdots\phi_{iT}}.$$

شایان ذکر است که در روش لگاریتم نسبت بخت وابستگی، در این مثال خاص، $\Psi_i = [\text{logit}(\phi_{i1}), \text{logit}(\phi_{i2})]$ و $\Omega_i = \log(w_{i12}) = \lambda_{i12}$ است.

هر یک از رهیافت‌هایی که در این بخش بیان شده‌اند، دارای محاسن و معایبی است. هنگامی که پارامترهای همبستگی به‌صورت توابعی از لگاریتم نسبت بخت شرطی بیان می‌شوند، پارامترهای همبستگی، مقید به میانگین حاشیه‌ای نیستند و فضای پارامترها برای $\Omega_i = [\log(\omega_{i12}), \ldots, \log(\omega_{i12\ldots T})]'$، به‌صورت R^{2^T-T-1} است؛ ولی تعداد پارامترها به T (تعداد پاسخ‌های تکرار شده برای هرآزمودنی) وابسته است. به عبارت دیگر می‌توان گفت این نوع مدل‌بندی، برای داده‌هایی با تعداد پاسخ‌های تکرار شده‌ی متفاوت برای آزمودنی‌های مختلف، قابل کاربرد نیستند. هنگامی که پارامترهای همبستگی بین پاسخ‌ها به‌صورت توابعی از لگاریتم نسبت بخت حاشیه‌ای بیان شوند، پارامترهای همبستگی آن، بر خلاف مدل‌هایی که از لگاریتم نسبت بخت شرطی برای همبستگی استفاده می‌کنند، به میانگین‌های حاشیه‌ای مقید می‌باشند؛ اما این قید به پیچیدگی قید در نظر گرفته‌شده برای همبستگی‌های حاشیه‌ای نیست. همچنین در این گونه مدل‌ها (لگاریتم نسبت بخت حاشیه‌ای)، بر خلاف مدل لگاریتم نسبت بخت شرطی، پارامترهای همبستگی به T وابسته نیستند؛ یعنی برای داده‌هایی با تعداد متفاوت از پاسخ‌های تکرار شده برای

آزمودنی‌های مختلف نیز قابل بیان است. همچنین مزیت این مدل نسبت به مدل لگاریتم نسبت بخت شرطی به‌عنوان پارامترهای همبستگی، این است که در مدل لگاریتم نسبت بخت حاشیه‌ای، قابلیت تولید شدن دوباره وجود دارد؛ در حالی که در مدل لگاریتم نسبت بخت شرطی، این انعطاف وجود ندارد. هنگامی که پارامترهای همبستگی بین پاسخ‌ها بر اساس نسبت وابسته بیان می‌شوند، پارامترهای همبستگی همانند لگاریتم نسبت بخت حاشیه‌ای، وابسته به T نیستند و به همین دلیل برای تعداد پاسخ‌های تکرار شده‌ی متفاوت برای آزمودنی‌های مختلف، این گونه مدل‌ها کاربرد دارند. همچنین این پارامترها (نسبت‌های وابسته) نیز به میانگین‌های حاشیه‌ای مقید می‌باشند، اما این قید به پیچیدگی قید همبستگی نیست.

لگاریتم نسبت بخت حاشیه‌ای

در این مدل‌بندی، در معادله‌ی (۱.٦) پارامترهای Ψ_i، عبارت‌هایی از احتمال‌های حاشیه‌ای هستند. به عبارت دیگر،

$$\psi_{it} = \text{logit}\{\Pr(Y_{it} = ۱)\},$$

و پارامترهای دوطرفه‌ی Ω_i، لگاریتم نسبت بخت حاشیه‌ای هستند. به‌عبارت دیگر،

$$\exp(w_{irs}) = \text{OR}(Y_{ir}, Y_{is}) = \frac{\Pr(Y_{ir}=۱, Y_{is}=۱)\Pr(Y_{ir}=۰, Y_{is}=۰)}{\Pr(Y_{ir}=۰, Y_{is}=۱)\Pr(Y_{ir}=۱, Y_{is}=۰)}.$$

پارامترهای سه‌طرفه Ω_i و بالاتر، به‌صورت توابعی از لگاریتم نسبت بخت حاشیه‌ای‌اند؛ به‌عنوان مثال

$$\exp(w_{irsl}) = \frac{\text{OR}(Y_{ir}, Y_{is}|Y_{il}=۱)}{\text{OR}(Y_{ir}, Y_{is}|Y_{il}=۰)}.$$

۲.۳.٦ مدل‌های اثرهای تصادفی

دومین رهیافت، مدل اثرهای تصادفی است که فرض همبستگی در میان پاسخ‌های تکرار شده را به‌وسیله‌ی ضرایب رگرسیونی متفاوت در آزمودنی‌های مختلف نشان می‌دهد (ویر، ۱۹۸۵). در مدل اثرهای تصادفی، پاسخ به‌عنوان تابعی از متغیرهای تبیینی با ضرایب

رگرسیونی‌ای فرض شده که از یک آزمودنی به آزمودنی دیگر متفاوت است. این تفاوت‌ها به‌خاطر عوامل اندازه‌گیری‌نشده‌ای است که منتج از عوامل طبیعی و ژنتیکی است. مثلاً رگرسیون خطی ساده برای رشد کودک را در نظر گیرید که ضرایب مدل، بیانگر وزن تولد و شاخص رشد است. واضح است که بچه‌ها در زمان تولد، دارای وزن و شاخص رشد متفاوت با یکدیگر به‌دلیل عوامل محیطی و ژنتیکی هستند. برای این مثال، یک مدل ضرایب تصادفی مناسب است، چرا که عوامل غیر قابل کنترل مانند عوامل محیطی و ژنتیکی وجود دارد که به‌عنوان یک متغیر تصادفی از یک آزمودنی به آزمودنی دیگر متفاوت است. اثر تصادفی می‌تواند به دو صورت در مدل ظاهر شود. نوع ساده‌ی این مدل عبارت است از

$$y_{it} = x'_{it}\beta_t + \tau_i + \varepsilon_{it},$$

که در آن x_{it} بردار متغیرهای تبیینی است، τ_i اثر تصادفی‌ای است که به‌صورت عرض از مبدأ تصادفی در مدل ظاهر شده است، و بردارهای ضرایب (β_t) ثابت ولی نامعلوم در نظر گرفته شده‌اند. می‌توان τ_i را با تابع توزیع یکسان و مستقل از هم با $E(\tau_i) = ۰$ و پارامتر نامعلوم $\mathrm{var}(\tau_i) = \sigma^۲_\tau$ در نظر گرفت.

مدل دیگر، مدلی است که در آن اثر تصادفی به‌صورت شیب و عرض از مبدأ تصادفی در مدل ظاهر می‌شود. در این مدل نه‌تنها عرض از مبدأ، بلکه برخی از ضرایب مدل، تصادفی هستند. مدل زیر، بیانگر این مسئله است:

$$y_{it} = x'_{it}\beta_t + z'_{it}\tau_i + \varepsilon_{it}, \tag{۲.۶}$$

که در آن z'_{it} زیربرداری از x'_{it} است و β_t برداری از پارامترهای ثابت ولی نامعلوم، و τ_i نیز یک اثر تصادفی چندمتغیره است. ضرایب رگرسیونی τ_i عموماً دارای توزیع نرمال با میانگین صفر و ماتریس کوواریانس Q فرض می‌شوند. حالت ساده‌ای از مدل (۲.۶) با در نظر گرفتن $\tau_i = (\tau_{i۱}, \tau_{i۲})$ و $\beta_t = \beta_{it}$، به‌صورت زیر بیان می‌شود:

$$E(Y_{it}|\tau_i) = \beta_۰ + \tau_{i۱} + (\tau_{i۲} + \beta_{۱t})x_{it},$$

که در آن، هم عرض از مبدأ و هم شیب، تصادفی فرض شده‌اند. مشخصات عمومی مدل‌های خطی تعمیم‌یافته‌ی اثرهای تصادفی به‌صورت زیر است:

آ) پاسخ‌های Y_{i1}، ...، Y_{iT_i} به‌شرط τ_i دوبه‌دو مستقل هستند و

$$f(y_{it}|\tau_i) = \exp[(y_{it}\theta_{it} - \psi(\theta_{it})/\phi + c(y_{ij}, \phi)].$$

ب) گشتاورهای شرطی $\mu_{it} = E(Y_{it}|\tau_i) = \psi'(\theta_{it})$ و $\nu_{it} = \text{var}(Y_{it}|\tau_i) = \psi''(\theta_{it})\phi$ در روابط زیر صادق هستند:

$$\theta_{it} = g(\mu_{it}) = x'_{it}\beta_t + z'_{it}\tau_i,$$

و

$$\nu_{it} = \nu(\mu_{it})\phi,$$

که در آن‌ها g و ν به‌ترتیب، توابع ربط و واریانس هستند و z_{it} زیربرداری از x_{it}ها است.

۳.۳.۶ مدل‌های انتقالی

مدل‌های انتقالی، مدل‌هایی هستند که در آن‌ها متغیر پاسخ در زمان t به متغیرهای پاسخ در زمان‌های $t-1$، ...، $t-q$ وابسته باشد، که $q < t - 1$ است. تحت یک مدل انتقالی، همبستگی بین پاسخ‌ها توسط تأثیر صریح مقادیر گذشته‌ی Y_{i1}، ...، Y_{it-1} روی Y_{it} مورد بررسی قرار می‌گیرد. پیامدهای گذشته به‌عنوان متغیرهای پیشگوی اضافی تلقی می‌شوند. مدل انتقالی با استفاده از مفهوم متغیر پنهان به‌صورت زیر بیان می‌شود:

$$Y_{it} = x'_{it}\beta_t + \varepsilon_{it},$$

که در آن:

$$\varepsilon_{it} = \gamma\varepsilon_{i,t-1} + z_{it},$$

که $\gamma = \exp(-\phi)$ است و z_{it}ها متغیرهای تصادفی نرمال با میانگین ۰ و واریانس σ_z^2 هستند، که $\sigma_z^2 = \sigma_\varepsilon^2(1 - \gamma^2)$. با جایگزین کردن $\varepsilon_{it} = Y_{it} - x'_{it}\beta_t$ در معادله‌ی دوم و تکرار این عمل، می‌توانیم به توزیع شرطی Y_{it} به‌شرط پاسخ گذشته ($Y_{i,t-1}$)، به‌صورت زیر دست یابیم:

$$Y_{it}|Y_{i,t-1} \sim \text{N}\{x'_{it}\beta_t + \gamma(Y_{i,t-1} - x'_{i,t-1}\beta_t), \sigma_z^2\}$$

حال اگر داشته باشیم

$$\varepsilon_{it} = \gamma_۱\varepsilon_{i,t-۱} + \cdots + \gamma_q\varepsilon_{i,t-q} + z_{it},$$

به توزیع شرطی Y_{it} به‌شرط پاسخ‌های گذشته، $Y_{i,t-۱}$، ...، $Y_{i,t-q}$، به‌صورت

$$Y_{it}|Y_{i,t-۱},\dots,Y_{i,t-q} \sim \mathrm{N}\left(x'_{it}\beta_t + \left[\sum_{j=۱}^{q}\gamma_j(Y_{i,t-j} - x'_{i,t-j}\beta_t)\right], \sigma_z^۲\right)$$

دست خواهیم یافت. چون پاسخ‌های ما گسسته‌اند، توزیع شرطی Y_{it} به‌شرط پاسخ‌های گذشته را می‌توان به‌صورت

$$\operatorname{logit}\Pr(Y_{it} = ۱|y_{i,t-۱},\dots,y_{i,t-q}) = x'_{it}\beta_t + \sum_{j=۱}^{q}\gamma_j y_{i,t-j}$$

مدل‌بندی کرد، که به مدل انتقالی خطی تعمیم‌یافته با تابع ربط لوجیت معروف است. توزیع شرطی Y_{it} به‌شرط تابعی از q پاسخ گذشته، برنولی است و به‌صورت زیر نمایش داده می‌شود:

$$f(y_{it}|y_{i,t-۱},\dots,y_{i,t-q},x'_{it}) = p_{it}^{y_{it}}(۱-p_{it})^{۱-y_{it}},$$

و

$$p_{it} = \frac{\exp(x'_{it}\beta_t + \sum_{j=۱}^{q}\gamma_j y_{i,t-j})}{۱+\exp(x'_{it}\beta_t + \sum_{j=۱}^{q}\gamma_j y_{i,t-j})}.$$

به‌عنوان مثال، در یک تحقیق پزشکی علاقه‌مندیم که بدانیم آیا علت بیماری آسم برای هر آزمودنی، به سابقه‌ی بیماری آزمودنی در دوره‌های قبل وابسته است یا خیر. برای بررسی این موضوع از مدل انتقالی استفاده می‌کنیم.

۴.۶ روش‌های براورد

۱.۴.۶ معادلات براوردگر تعمیم‌یافته برای پاسخ‌های دودویی

این رهیافت توسط لیانگ و زیگر (۱۹۸۶) و پرنتیس (۱۹۸۸) پیشنهاد شده است. رهیافت معادلات براوردگر تعمیم یافته[۱] (GEE)، رهیافت واحدی برای تحلیل برامدهای

[۱] Generalized Estimating Equations

گوناگون پیوسته و گسسته پیشنهاد می‌کند. در این میان، دسته‌ای از معادلات براوردگر تعمیم‌یافته برای پارامترهای مدل رگرسیونی پیشنهاد شده است. در این رهیافت، تنها فرض‌هایی که اختیار می‌شوند، فرض در باره‌ی امید ریاضی حاشیه‌ای مرتبه‌ی اول و دوم پاسخ‌ها است و در باره‌ی توزیع کامل آن‌ها فرضی اختیار نمی‌شود.

این رهیافت برای پارامترهای مدل رگرسیونی و واریانس آن‌ها براوردگرهای سازگاری ارائه می‌کند. در اصل، GEE حالت چندمتغیره‌ی شبه‌درست‌نمایی است که توسط ودربرن (۱۹۷۴) بیان شده است. در این رهیافت، تابعی از امید ریاضی حاشیه‌ای متغیر وابسته، به‌صورت تابعی خطی از متغیرهای کمکی مشخص می‌شود. همچنین فرض می‌شود که واریانس، تابعی معلوم از میانگین حاشیه‌ای است. به‌علاوه، ماتریس همبستگی عملی برای مشاهدات هر آزمودنی، معین می‌شود. معادلات براوردگر تعمیم‌یافته‌ی لیانگ و زیگر (۱۹۸۶) برای براورد کردن بردار $(p \times 1)$—بعدی از پارامترهای مدل رگرسیونی، β، تعمیم معادلات براوردگر به داده‌های همبسته است که به‌صورت زیر تعریف می‌شود:

$$S(\beta) = \sum_{i=1}^{n} D_i{}' V_i^{-1}(Y_i - \mu_i(\beta)) = 0, \qquad (۳.۶)$$

که در آن $D_i = \frac{\partial \mu_i}{\partial \beta}$ است. همچنین $g(\mu_{it}) = x'_{it}\beta$ است که در آن، g تابع ربط و x_{it} برداری p—بعدی از متغیرهای کمکی است. ماتریس $(p \times T_i)$—بعدی از مشتقات جزئی میانگین نسبت به پارامترهای مدل رگرسیونی برای iامین فرد، به‌صورت زیر مشخص می‌شود:

$$D_i' = \frac{\partial \mu_i'}{\partial \beta} = \begin{bmatrix} \frac{x_{i11}}{g'(\mu_{i1})} & \cdots & \frac{x_{iT_i1}}{g'(\mu_{iT_i})} \\ \vdots & & \vdots \\ \frac{x_{i1p}}{g'(\mu_{i1})} & \cdots & \frac{x_{iT_ip}}{g'(\mu_{iT_i})} \end{bmatrix}_{p\times T_i}$$

ماتریس واریانس— کوواریانس V_i می‌تواند به‌صورت زیر نمایش داده شود:

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

که ماتریس واریانس— کوواریانس Y_iها است و در آن، A_i ماتریس قطری $(T_i \times T_i)$—بعدی با t امین عنصر قطری $v(\mu_{it})$ است. همچنین می‌توان V_i را به‌صورت

$$V_i = \phi A_i^{\frac{1}{2}} W_i^{-\frac{1}{2}} R_i(\alpha) W_i^{-\frac{1}{2}} A_i^{\frac{1}{2}}$$

تعریف کرد که در آن، W_i ماتریس قطری $(T_i \times T_i)$—بعدی با عناصر W_{it} به‌عنوان tامین عنصر قطری است که معرف وزن مربوط به iامین فرد در t امین زمان است. اگر وزن همه‌ی افراد در زمان‌های مختلف، ۱ باشد، W_{it} برای همه‌ی افراد در زمان t، برابر با ۱ است (مدل لیانگ و زیگر، ۱۹۸۶). ساختارهای مختلفی برای ماتریس همبستگی عملی وجود دارد که از آن جمله می‌توان به موارد زیر اشاره کرد: نوع اول، ناهمبستگی است که در آن $R = I$. در این حالت، معادلات براوردگر تعمیم‌یافته به معادلات براوردگر مستقل تبدیل می‌شوند. نوع دیگر، همبستگی تبادل‌پذیر است که در آن، مشاهدات هر آزمودنی، همبستگی یکسانی دارند و به‌صورت زیر نمایش داده می‌شود:

$$R = \operatorname{corr}(Y_{it}, Y_{it'}) = \begin{cases} ۱, & t = t' \\ \alpha, & t \neq t' \end{cases}$$

همبستگی اتورگرسیو دارای ساختار همبستگی $R = \alpha^{|t-t'|}$ است. یک راه حل دیگر برای در نظر گرفتن وابستگی زمانی، این است که این همبستگی فقط برای برخی از واحدها وجود داشته باشد. این نوع همبستگی به همبستگی مانا معروف است و به‌صورت زیر تعریف می‌شود:

$$R_{uv} = \begin{cases} \alpha_{|u-v|}, & |u-v| \leq k \\ ۰, & \text{در غیر این صورت} \end{cases}$$

همبستگی نامانا نوعی دیگر از ساختار همبستگی است. فرق ساختار همبستگی نامانا با همبستگی مانا در این است که در ساختار همبستگی نامانا، همبستگی ثابت برای همه‌ی عناصر زیر قطر اصلی فرض نشده است. یکی دیگر از انواع همبستگی، همبستگی غیر ساختاری است که به‌صورت زیر معرفی می‌شود:

$$\operatorname{corr}(Y_{it}, Y_{it'}) = \begin{cases} ۱, & t = t' \\ \alpha_{tt'}, & t \neq t' \end{cases}$$

اگر اطلاعات در باره‌ی ساختار ماتریس همبستگی از منابع دیگر موجود باشد، ماتریس همبستگی ثابت پیشنهاد می‌شود. در این روش، ماتریس همبستگی عملی در هر مرحله براورد نمی‌شود، بلکه ماتریس همبستگی همان‌طور که مشخص شده است، در نظر گرفته می‌شود.

اشاره به نکات زیر در باره‌ی معادله‌ی (۳.۶) لازم است. اول این‌که اگر در معادله‌ی

(۳.۶) به‌ازای هر i، که $i = 1, \dots, n$، داشته باشیم $T_i = 1$، این معادله به معادله‌ی شبه‌درست‌نمایی کاهش می‌یابد. دوم این‌که برای i امین آزمودنی، $U_i(\beta, \alpha) = D_i' V_i^{-1}(Y_i - \mu_i(\beta))$ شبیه تابعِ به دست آمده از رهیافت شبه‌درست‌نمایی است، به‌جز آن‌که V_iها در این‌جا تابعی از α و β می‌باشند و نه این‌که صرفاً تابعی از β باشند.

معادله‌ی (۳.۶) می‌تواند به‌صورت تابعی صرفاً از β بیان شود، به این‌صورت که نخست α در فرمول V_i به‌وسیله‌ی $\hat{\alpha}(Y, \beta, \phi)$ جایگذاری می‌شود (براوردگر $n^{\frac{1}{2}}$—سازگار برای α هنگامی که β و ϕ معلوم هستند؛ به این معنا که $n^{\frac{1}{2}}(\hat{\alpha} - \alpha) = o_p(1)$ است. به عبارت دیگر به‌ازای هر $\varepsilon > 0$، هنگامی که $n^{\frac{1}{2}} \to \infty$، داریم $\Pr[n^{\frac{1}{2}}(\hat{\alpha} - \alpha) < \varepsilon] = 0$). به‌جز برای حالت‌های خاصی از R و $\hat{\alpha}$، پارامتر مقیاس ϕ می‌تواند در معادله‌ی (۳.۶) باقی بماند. برای کامل شدن فرایند، ϕ به‌وسیله‌ی $\hat{\phi}(Y, \beta)$ (براوردگر $n^{\frac{1}{2}}$—سازگار برای ϕ، زمانی که β معلوم باشد) جایگزین می‌شود. در نهایت، معادله‌ی (۳.۶) به‌صورت زیر در خواهد آمد:

$$\sum_{i=1}^{n} U_i[\beta, \hat{\alpha}\{\beta, \hat{\phi}(\beta)\}] = 0 \qquad (۴.۶)$$

و $\hat{\beta}_G$ که نشان‌دهنده‌ی براورد β تحت رهیافت GEE است، جواب معادله‌ی (۴.۶) خواهد بود.

قضیه‌ی زیر که توسط لیانگ و زیگر (۱۹۸۶) بیان و ثابت شده است، خواص نمونه‌های بزرگ $\hat{\beta}_G$ را بیان می‌کند.

قضیه‌ی ۱.۶: تحت شرایط زیر

(i) به‌شرطی که β و ϕ معلوم باشند، $\hat{\alpha}$ براوردگر $n^{\frac{1}{2}}$—سازگار است؛

(ii) به‌شرطی که β معلوم باشد، $\hat{\phi}$ براوردگر $n^{\frac{1}{2}}$—سازگار است؛

(iii) $\left|\frac{\partial \hat{\alpha}(\beta, \phi)}{\partial \phi}\right| \leq H(Y, \beta)$ که $O_p(1)$ است؛

$n^{\frac{1}{2}}(\hat{\beta}_G - \beta)$ به‌طور مجانبی دارای توزیع چندمتغیره‌ی گاوسی با میانگین صفر و ماتریس کوواریانس V_G است که به‌صورت زیر تعریف می‌شود:

$$V_G = \lim_{n \to \infty} n\left(\sum_{i=1}^{n} D_i' V_i^{-1} D_i\right)^{-1}$$

$$\times\left\{\sum_{i=1}^{n} D_i' V_i^{-1} \mathrm{cov}(Y_i) V_i^{-1} D_i\right\}(\sum_{i=1}^{n} D_i' V_i^{-1} D_i)^{-1} = \lim_{n\to\infty} nV_1^{-1} V_0 V_1^{-1}$$

براورد واریانس $\hat{\beta}_G$ (یعنی $\hat{V}_G$) می‌تواند به‌وسیله‌ی جایگذاری $\mathrm{cov}(Y_i)$ با $(Y_i - \mu_i(\beta))(Y_i - \mu_i(\beta))'$، و نیز جایگذاری α و β و ϕ و با براوردگرهای آن‌ها در V_G حاصل شود. توجه کنید که براورد کوواریانس مجانبی $\hat{\beta}_G$ به انتخاب براوردگرهای $n^{\frac{1}{2}}$—سازگار α و ϕ وابسته نیست. این نتایج، مشابه شبه‌درست‌نمایی برای واریانس پارامترهای مدل رگرسیونی است که وابسته به انتخاب براوردگر ϕ نیست.

برای به دست آوردن $\hat{\beta}_G$، از الگوریتم برازش مدل با استفاده از معادلات براوردگر تعمیم‌یافته به‌صورت زیر استفاده می‌شود:

۱— مقادیر اولیه‌ی $\beta^{(0)}$ را با یک مدل خطی تعمیم‌یافته‌ی معمولی، با فرض استقلال، محاسبه می‌کنیم.

۲— ماتریس همبستگی عملی بر اساس مانده‌های استاندارد شده و در نظر گرفتن یک ساختار برای R محاسبه می‌شود.

۳— براورد کوواریانس با استفاده از رابطه‌ی زیر محاسبه می‌شود:

$$V_i = \phi A_i^{\frac{1}{2}} W_i^{-\frac{1}{2}} \hat{R}_i(\alpha) W_i^{-\frac{1}{2}} A_i^{\frac{1}{2}}.$$

۴— مقدار β را با استفاده از رابطه‌ی زیر، بهنگام می‌کنیم:

$$\beta_{r+1} = \beta_r + [\sum_{i=1}^{n} \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta}]^{-1} [\sum_{i=1}^{n} \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} (Y_i - \mu_i)]$$

۵— مراحل ۲ تا ۴ را تا رسیدن به همگرایی، تکرار می‌کنیم.

اکنون به براورد کردن پارامترهای α و ϕ می‌پردازیم. در یک تکرار مشخص، پارامتر همبستگی α و پارامتر مقیاس ϕ می‌توانند از مانده‌های پی‌یرسون براورد شوند که به‌صورت زیر معرفی می‌شوند:

$$\hat{r}_{it} = \frac{y_{it} - \hat{\mu}_{it}}{\sqrt{v(\hat{\mu}_{it})}},$$

که $\hat{\mu}_{it}$ به مقادیر β وابسته است. براورد کردن ϕ به‌عنوان پارامتر پراکندگی، به‌صورت زیر انجام می‌شود:

$$\hat{\phi}^{-1} = \frac{1}{N-p}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\hat{r}_{it}^{2},$$

که این کمّیت مشابه حالت طولی آماره‌های پی‌یرسون است (ودربرن، ۱۹۷۴، و مکولا، ۱۹۸۳) که در آن $N = \sum_{i=1}^{n} T_i$ تعداد کل مقادیر اندازه‌گیری‌شده است و p تعداد پارامترهای مدل رگرسیونی است. نکته‌ی قابل توجه این است که براوردها در این روش به‌طور کلّی بر اساس تابع درست‌نمایی به دست نمی‌آیند.

برای به دست آوردن براوردی سازگار برای α باید بر روی n آزمودنی تکیه شود. براورد α وابسته به انتخاب $R(\alpha)$ است. روش عمومی این است که α به‌وسیله‌ی تابع زیر براورد شود:

$$\hat{R}_{jt} = \frac{\sum_{i=1}^{n}\hat{r}_{ij}r_{it}}{N-p}.$$

پس از پیدا کردن $\hat{\alpha}$ و $\hat{\phi}$، آن‌ها را در معادله‌ی (۴.۶) قرار داده، $\hat{\beta}_G$ را به دست می‌آوریم و به همین ترتیب ادامه می‌دهیم تا همگرایی حاصل شود. همان‌طور که گفته شد، براورد α وابسته به انتخاب $R(\alpha)$ است. به مثال زیر در این زمینه توجه کنید.

مثال: فرض کنید $\operatorname{corr}(Y_{it}, Y_{it'}) = \alpha^{|t-t'|}$. برای یک y_{it} که دارای توزیع گاوسی است، این ساختار همبستگی زمان پیوسته، شبیه فرایند اتورگرسیو مرتبه‌ی اول $AR(1)$ است (فلر، ۱۹۷۱، ص. ۸۹). چون تحت این مدل ، رابطه‌ی

$$E(\hat{r}_{it}r_{it'}) \approx \alpha^{|t-t'|}$$

برقرار است، می‌توان α را از طریق شیب خط رگرسیونی $\log(\hat{r}_{it}\hat{r}_{it'})$ روی $\log|t-t'|$ براورد کرد.

مثال: اگر برای همه‌ی tها $(t \neq t')$ داشته باشیم $\operatorname{corr}(Y_{it}, Y_{it'}) = \alpha$ (به این ساختار همبستگی، ساختار قابل تعویض گفته می‌شود)، با فرض معلوم بودن ϕ، می‌توان α را به‌صورت زیر براورد کرد:

$$\hat{\alpha} \quad = \quad \frac{\phi\sum_{i=1}^{n}\sum_{t>t'}\hat{r}_{it}\hat{r}_{it'}}{\sum_{i=1}^{n}\frac{T_i(T_i-1)}{2}-p}$$

$$= \frac{\phi \sum_{i=۱}^{n} \sum_{t>t'} \hat{r}_{it}\hat{r}_{it'}}{\sum_{i=۱}^{m} \frac{T(T-۱)}{۲} - p}$$

فیتس‌موریس و دیگران (۱۹۹۳) برای براورد α، مجموعه‌ای دیگر از معادلات براورد گشتاوری، شبیه معادله‌ی (۳.۶) بیان کردند. این مجموعه، به‌صورت زیر است:

$$U(\alpha) = \sum_{i=۱}^{n} A_i' B_i^{-۱}(r_i - \rho_i(\alpha)) = ۰, \qquad (۵.۶)$$

که در آن $\rho_i(\alpha) = (\rho_{i۱۲}, \ldots, \rho_{iT-۱T})'$ و $A_i = \frac{\partial \rho_i(\alpha)}{\partial \alpha'}$ است و B_iها ماتریس کوواریانس عملی r_iها هستند. r_iها نیز همبستگی‌های تجربی هستند؛ یعنی

$$r_{ist} = \frac{(Y_{is} - \mu_{is})(Y_{it} - \mu_{it})}{[\mu_{is}(۱ - \mu_{is})\mu_{it}(۱ - \mu_{it})]^{\frac{۱}{۲}}},$$

که در آن $E(r_{ist}) = \rho_{ist}$.

B_i را می‌توان به‌صورت یک ماتریس همانی با بُعد n در نظر گرفت. در بسیاری از حالت‌ها، این فرض برای B_i، منجر به روش‌های غیر تکراری ساده برای محاسبه‌ی α می‌شود. به‌عنوان مثال، اگر همبستگی‌های جفتی به‌صورت $R_i(\alpha) = \rho$ فرض شوند، $\hat{\rho} = \frac{\sum_{i=۱}^{n} r_i}{n}$ است.

علاوه بر این، نیویی (۱۹۹۰) نشان داد که انتخاب براوردگر α، تأثیری بر کارایی مجانبی براوردگر β ندارد. همچنین فیتس‌موریس و دیگران (۱۹۹۳) براورد واریانس مجانبی $\hat{\beta}_G$ را به‌صورت زیر در نظر گرفتند:

$$\hat{V}_G^* = \left(\sum_{i=۱}^{n} \hat{D}_i' \hat{V}_i^{-۱} \hat{D}_i \right)^{-۱}.$$

آنان بیان کردند که اگر داشته باشیم $V_i = \mathrm{cov}(Y_i)$، یعنی زمان وابستگی به‌درستی مشخص شده باشد، $\hat{V}_G^*$ براوردگری سازگار است. اگر همبستگی عملی $R_i(\alpha)$ به‌درستی مشخص نشده باشد، $\hat{V}_G^*$ براوردگر سازگاری نخواهد بود؛ در حالی که حتی اگر کوواریانس عملی (V_i) برابر با $\mathrm{cov}(Y_i)$ نباشد، $\hat{V}_G$ در قضیه‌ی ۱.۶ سازگار است. به همین دلیل به $\hat{V}_G$ براوردگر استوار گفته می‌شود. در نتیجه رهیافت GEE دارای دو خاصیت جالب است:

آ) براوردی سازگار برای β ارائه می‌کند که فقط احتیاج به مدل‌بندی میانگین $(\mu_i(\beta))$ دارد. بنا بر این صرف نظر از این‌که همبستگی عملی به‌درستی مشخص شده باشد،

براوردی سازگار از ضرایب رگرسیونی به دست می‌دهد.

ب) براوردهای استوار، واریانس را هنگامی به دست می‌دهند که همبستگی عملی به‌درستی مشخص شده باشد.

۲.۴.۶ بسط درجه‌ی دوم GEE

این مدل که توسط لیانگ و زیگر (۱۹۸۶) معرفی شده، معروف به روش GEE2 است. در این روش، براورد پاسخ Y_i با حاصل‌ضرب‌های دوبه‌دوی آن‌ها، به‌صورت یک بردار پاسخ واحد با $Z_i = (Y_i', W_i')'$ مورد بررسی قرار می‌گیرد که دارای $T + \frac{T(T-1)}{2}$ درایه است. در این روش $(\hat{\beta}, \hat{\alpha})$ از رابطه‌ی زیر به دست می‌آید:

$$\begin{aligned} U(\delta) &= \sum_{i=1}^{n} \frac{\partial(\mu_i, \eta_i)}{\partial \delta} cov^{-1} \begin{pmatrix} Y_i \\ W_i \end{pmatrix} \cdot \begin{pmatrix} Y_i - \mu_i \\ W_i - \eta_i \end{pmatrix} \\ &= \sum_{i=1}^{n} C_i B_i A_i, \end{aligned} \tag{۶.۶}$$

که در آن $Y_i = (Y_{i1}, \ldots, Y_{iT})'$ و $\omega_i = (Y_{i1}Y_{21}, \ldots, Y_{iT-1}Y_{iT}, \ldots, Y_{i1}Y_{i2}\ldots Y_{iT})'$ و $\delta = (\beta, \alpha)$ و $\eta_i = E(\omega_i)$ است. همچنین طبق قضیه‌ی ۱.۶، $n^{\frac{1}{2}}(\hat{\delta} - \delta)$ دارای توزیع نرمال با میانگین صفر و ماتریس کوواریانس براورد شده‌ی زیر است:

$$\begin{aligned} \hat{V}_{\delta} &= \left(\sum_{i=1}^{n} C_i B_i C_i'\right)^{-1} \left(\sum_{i=1}^{n} C_i B_i \begin{pmatrix} Y_i - \mu_i \\ \\ W_i - \eta_i \end{pmatrix} \begin{pmatrix} Y_i - \mu_i \\ \\ W_i - \eta_i \end{pmatrix}' B_i C_i'\right) \\ &\quad \times \left(\sum_{i=1}^{n} C_i B_i C_i'\right)^{-1}. \end{aligned} \tag{۷.۶}$$

حال اگر دو معادله‌ی (۳.۶) و (۴.۶) از GEE را در یک معادله قرار دهیم، خواهیم داشت:

$$U(\delta) = \sum_{i=1}^{n} C_i B_i^* A_i, \tag{۸.۶}$$

که B_i^* ماتریسی به‌صورت زیر است:

$$B_i^* = \begin{pmatrix} \mathrm{cov}(Y_i) & 0 \\ 0 & \mathrm{cov}(W_i) \end{pmatrix}.$$

تفاوت GEE با GEE2 این است که در روش GEE، کوواریانس بین Y_i و W_i در نظر گرفته نمی‌شود؛ در حالی که در روش GEE2، این کوواریانس در محاسبات به حساب می‌آید.
ضعف کلّی روش GEE2 این است که اگر کوواریانس Y_iها به‌درستی مشخص نشود، براوردگرهای β سازگار نیستند؛ در حالی که در GEE این براوردگرها حتی هنگامی که ماتریس کوواریانس به‌درستی مشخص نشده باشد، براوردگرهای سازگاری هستند.
همچنین در معادله‌ی (٦.٦) حتی اگر $\operatorname{cov}(Y_i)$ها به‌درستی مشخص شده باشند، نیاز به گشتاورهای مراتب بالاتر Y_i دارد که محاسبه‌ی آن‌ها بسیار وقت‌گیر است.

مزیت روش GEE2 نسبت به GEE این است که براوردگرها در این روش، دارای کارایی زیادی هستند؛ زیرا همان‌طور که گفته شد، در معادله‌ی (٦.٦) همبستگی بین مؤلفه‌های Y_i و W_i، برابر با صفر در نظر گرفته می‌شود، ولی در GEE2 این کوواریانس، مخالف صفر در نظر گرفته می‌شود. به این ترتیب در GEE از اطلاعات کمتری استفاده می‌شود و در نتیجه دارای کارایی کمتری است.

مثال: در این مثال، رهیافت GEE برای تحلیل داده‌های طولی با دو پاسخ برای هر آزمودنی، مورد بحث قرار می‌گیرد.
همان‌طور که می‌دانیم، معادله‌ی رهیافت GEE به‌صورت زیر خواهد بود:

$$\sum_{i=1}^{n}\left(\frac{\partial \mu_i}{\partial \beta}\right)' \operatorname{var}(Y_i)^{-1}(Y_i-\mu_i)=0$$

برای داده‌های طولی با دو پاسخ برای هر آزمودنی، $\mu_i=\begin{bmatrix}\mu_{i1}\\ \mu_{i2}\end{bmatrix}$ است که در آن $\mu_{it}=\frac{\exp(x'_{it}\beta)}{1+\exp(x'_{it}\beta)}$ و

$$\operatorname{var}(Y_i)=\begin{bmatrix}\mu_{i1}(1-\mu_{i1}) & \operatorname{cov}(Y_{i1},Y_{i2})\\ \operatorname{cov}(Y_{i1},Y_{i2}) & \mu_{i1}(1-\mu_{i1}),\end{bmatrix},$$

که $\operatorname{cov}(Y_{i1},Y_{i2})=\rho\sqrt{\mu_{i1}(1-\mu_{i1})\mu_{i2}(1-\mu_{i2})}$ خواهد بود. در نتیجه

$$\operatorname{var}(Y_i)^{-1}=\frac{1}{\mu_{i1}(1-\mu_{i1})\mu_{i2}(1-\mu_{i2})(1-\rho_i^2)}$$

$$\times\begin{bmatrix}\mu_{i2}(1-\mu_{i2}) & -\rho\sqrt{\mu_{i1}(1-\mu_{i1})\mu_{i2}(1-\mu_{i2})}\\ -\rho\sqrt{\mu_{i1}(1-\mu_{i1})\mu_{i2}(1-\mu_{i2})} & \mu_{i1}(1-\mu_{i1})\end{bmatrix}.$$

اگر $\rho_i = \alpha$ (به‌ازای هر i) فرض شود، معادله‌ی براورد GEE به‌صورت زیر است:

$$\sum_{i=1}^{n} \left(\frac{\partial \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix}}{\partial \beta} \right)' \operatorname{var}(Y_i)^{-1} \begin{bmatrix} y_{i1} - \mu_{i1} \\ y_{i2} - \mu_{i2} \end{bmatrix} = 0,$$

و با توجه به این مطلب که

$$\frac{\partial \mu_{it}}{\partial \beta} = x_{it}\mu_{it}(1 - \mu_{it})$$

است، در معادله‌ی رهیافت GEE خواهیم داشت:

$$\sum_{i=1}^{n} \left\{ \frac{x_{i1}(y_{i1} - \mu_{i1})}{1-\alpha^2} - \frac{\alpha x_{i2}(y_{i1} - \mu_{i1})}{1-\alpha^2} \sqrt{\frac{\mu_{i2}(1-\mu_{i2})}{\mu_{i1}(1-\mu_{i1})}} \right.$$
$$\left. \times \frac{x_{i2}(y_{i2} - \mu_{i2})}{1-\alpha^2} - \frac{\alpha x_{i1}(y_{i2} - \mu_{i2})}{1-\alpha^2} \sqrt{\frac{\mu_{i1}(1-\mu_{i1})}{\mu_{i2}(1-\mu_{i2})}} \right\} = 0,$$

که همان‌طور که قبلاً بیان شد، براورد مؤلفه‌های β جواب این معادله‌ها است.

۳.۴.۶ معادلات درست‌نمایی و براوردگرهای ماکسیمم درست‌نمایی

در این بخش، معادلات درست‌نمایی و براوردهای آن برای مدل‌های معرفی‌شده در بخش ۳.۶ ارائه می‌شود. در ابتدا به مدل‌های حاشیه‌ای بر اساس درست‌نمایی می‌پردازیم. در همه‌ی این مدل‌ها، توزیع مشترک پاسخ‌ها با در نظر گرفتن فرض‌هایی، به‌طور کامل مشخص می‌شود. یکی از این فرض‌ها، فرض در باره‌ی ارتباط‌های پاسخ‌ها است. در این بخش، رهیافت‌های حاشیه‌ای که معیارهای همبستگی آن‌ها مقید به احتمال‌های حاشیه‌ای نیستند (لگاریتم نسبت بخت شرطی، نسبت وابسته و لگاریتم نسبت بخت حاشیه‌ای)، بیان می‌شوند. توجه داشته باشید که برای نوشتن تابع درست‌نمایی ، نخست باید شیوه‌ی در نظر گرفتن همبستگی‌ها تشریح شود. همچنین در همه‌ی این مدل‌ها ارتباط پاسخ‌ها با متغیرهای کمکی به‌صورت $\operatorname{logit}(\mu_{it}) = x'_{it}\beta_t$ و واریانس پاسخ‌ها به‌صورت $\operatorname{var}(Y_{it}) = \mu_{it}(1 - \mu_{it})$ بیان می‌شود. تنها تفاوت آن‌ها در معیار همبستگی پاسخ‌ها است.

۱.۳.۴.۶ مدل حاشیه‌ای با پارامترهای همبستگی به‌صورت توابعی از لگاریتم نسبت بخت شرطی

در این مدل‌بندی که توسط فیتس‌موریس و لرد (۱۹۹۳) پیشنهاد شده است، در معادله‌ی (۱.۶) همبستگی پاسخ‌ها برای هر آزمودنی بر اساس لگاریتم نسبت بخت شرطی و توابعی از آن‌ها قابل بیان خواهد شد.

در معادله‌ی (۱.۶) پارامترهای Ψ_i عبارت‌هایی از احتمال‌های شرطی به‌صورت $\psi_{ij} = \text{logit}\{\Pr(Y_{ij} = 1 | Y_{is} = 0, s \neq j)\}$ هستند و پارامترهای Ω_i عبارت‌هایی از لگاریتم نسبت بخت شرطی‌اند؛ یعنی:

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

با این نوع مدل‌بندی، درجه‌های مختلف وابستگی در میان Y_{it}ها قابل بیان است. به‌عنوان مثال، اگر همه‌ی پارامترهای همبستگی دوطرفه و بیش‌تر، صفر باشند، مدل استقلال به دست خواهد آمد، یا اگر همه‌ی درایه‌های Ω_i مخالف صفر باشند، مدل اشباع‌شده به دست می‌آید. بین این دو حالت، مدل‌های مختلفی از وابستگی را می‌توانیم در نظر بگیریم. یک حالت ویژه از این مدل‌ها، صورت خانواده‌ی نمایی درجه‌ی دوم یا مدل جفتی است که به‌وسیله‌ی صفر قرار دادن پارامترهای همبستگی سه‌طرفه و بیش‌تر حاصل می‌شود (ژائو و پرنتیک، ۱۹۹۰).

ژائو و پرنتیس (۱۹۹۰) تبدیلی یک‌به‌یک از (Ψ_i, Ω_i) به پارامترهای گشتاوری (μ_i, Λ_i) انجام دادند، که Λ_i برداری از همبستگی‌های حاشیه‌ای است. سپس این پارامترهای گشتاوری به‌صورت توابعی از متغیرهای کمکی به‌وسیله‌ی تابع‌های ربط معین، مدل‌بندی می‌شوند.

۲.۳.۴.۶ مدل حاشیه‌ای با پارامترهای همبستگی به‌صورت توابعی از نسبت وابسته

این مدل حاشیه‌ای توسط اکهولم و دیگران (۱۹۹۵) پیشنهاد شده است. در این روش،

استفاده از پارامتر میانگین (مؤلفه‌هایی که احتمال موفقیت مشترک همه‌ی مراتب را در بر دارد) برای تحلیل رگرسیون پاسخ دودویی چندمتغیره پیشنهاد شده است که همبستگی را با استفاده از نسبت‌های وابسته (که عبارت‌هایی از پارامتر میانگین‌اند) بیان می‌کند.
همان‌گونه که بیان شد، W_i در معادله‌ی (۱.۶)، بردار حاصل‌ضرب‌های دوتایی و بیش‌تر است که به‌صورت زیر تعریف می‌شود:

$$\omega_i = (Y_{i۱}, \ldots, Y_{iT}, Y_{i۱}Y_{i۲}, \ldots, Y_{iT-۱}Y_{iT}, \ldots, Y_{i۱}Y_{i۲}\ldots Y_{iT})'$$

و به بردار مقادیر مورد انتظار ω_i، پارامتر میانگین (ϕ_i) می‌گویند (براندولف نیلسن و کاکس، ۱۹۹۴)؛ یعنی:

$$\phi_i = (\phi_{i۱}, \ldots, \phi_{iT}, \phi_{i۱۲}, \ldots, \phi_{i(T-۱)T}, \ldots, \phi_{i۱۲\ldots T})'.$$

مؤلفه‌های ϕ_i، که احتمال‌های موفقیت مشترک حاشیه‌ای هستند، به‌صورت زیر بیان می‌شوند:

$$\begin{aligned}
\phi_{i۱} &= \Pr(Y_{i۱} = ۱),\\
\phi_{i۱۲} &= \Pr(Y_{i۱} = ۱, Y_{i۲} = ۱),\\
&\vdots\\
\phi_{i۱۲\ldots T} &= \Pr(Y_{i۱} = ۱, \ldots, Y_{iT} = ۱).
\end{aligned}$$

این پارامترها بر خلاف لگاریتم نسبت بخت‌های شرطی، وابسته به n نیستند؛ به‌عنوان مثال، $\phi_{i۱۲۳۵}$ دارای تفسیر یکسانی برای همه‌ی $n \geq ۵$ها است. اما لگاریتم نسبت بخت شرطی، مانند $\omega_{i۱۲}$ به‌شرط $Y_{i۳} = \cdots = Y_{iT} = ۰$ محاسبه می‌شود و همان‌طور که دیده می‌شود، به n وابسته است.
اکنون نشان می‌دهیم که احتمال‌های آزمودنی iام به‌صورت زیر بر حسب ϕ_i به دست می‌آیند. شمارنده‌ی تصادفی $I_{y_i}(Y_i)$ را به‌صورت زیر تعریف می‌کنیم:

$$I_y(Y_i) = \begin{cases} ۱, & Y_i = y_i \\ ۰, & Y_i \neq y_i \end{cases}$$

دو عبارت جبری مختلف برای $I_y(Y)$ وجود خواهد داشت:
۱) به‌صورت حاصل‌ضرب T عامل است. عامل‌ها به‌صورتی تعریف می‌شوند که اگر

$Y_{it}=1$باشد، Y_{it}ها به‌عنوان عامل در نظر گرفته شوند و اگر $Y_{it}=0$ باشد، $(1-Y_{it})$ها به‌عنوان عامل در نظر گرفته شوند.

۲) به‌صورت مجموعی از حاصل‌ضرب Y_{it} ها، بعد از بسط همه‌ی عامل‌های $(1-Y_{it})$ و ضرب عناصر موجود در یکدیگر، در نظر گرفته می‌شود.

به‌عنوان مثال، اگر $y_i=(0,1,1,0,1,1)$ را در نظر بگیریم، خواهیم داشت:

$$\begin{aligned}I_{y_i}(Y_i) &= (1-Y_{i1})Y_{i2}Y_{i3}(1-Y_{i4})Y_{i5}Y_{i6}\\ &= Y_{i2}Y_{i3}Y_{i5}Y_{i6}-Y_{i1}Y_{i2}Y_{i3}Y_{i5}Y_{i6}-Y_{i2}Y_{i3}Y_{i4}Y_{i5}Y_{i6}+Y_{i1}Y_{i2}Y_{i3}Y_{i4}Y_{i5}Y_{i6}.\end{aligned}$$

واضح است که $\Pr(Y_i=y_i)=E\{I_{y_i}(Y_i)\}$؛ یعنی:

$$\begin{aligned}\Pr(Y_i=y_i) &= E(Y_{i2}Y_{i3}Y_{i5}Y_{i6}-Y_{i1}Y_{i2}Y_{i3}Y_{i5}Y_{i6}\\ &\quad -Y_{i2}Y_{i3}Y_{i4}Y_{i5}Y_{i6}+Y_{i1}Y_{i2}Y_{i3}Y_{i4}Y_{i5}Y_{i6})\\ &= \phi_{i2356}-\phi_{i12356}-\phi_{i23456}+\phi_{i123456}.\end{aligned}$$

در نتیجه $\Pr(Y_i=y_i)$ می‌تواند بر حسب عبارتی‌هایی از مؤلفه‌های ϕ بیان شود. در این مدل‌بندی، ارتباط و همبستگی با نسبت وابسته، که عبارت‌هایی از پارامتر میانگین (ϕ_i) می‌باشند، مدل‌بندی می‌شود.

پاسخ دودویی دومتغیره برای آزمودنی iام، (Y_{i1},Y_{i2})، با پارامتر میانگین $\phi_i=(\phi_{i1},\phi_{i2},\phi_{i12})'$ را در نظر بگیرید. نسبت وابسته، ω_{i12}، و لگاریتم نسبت وابسته، λ_{i12}، به‌صورت زیر تعریف می‌شوند:

$$\omega_{i12}=\frac{\phi_{i12}}{\phi_{i1}\phi_{i2}},\qquad \lambda_{i12}=\log(\omega_{i12}), \tag{۶.۹}$$

که در آن‌ها $100\times(\omega_{i12}-1)$ درصد بزرگی احتمال موفقیت توأم Y_{i1} و Y_{i2} در مقایسه با فرض استقلال را نشان می‌دهد.

هنگامی که $\omega_{i12}=1$ باشد، $Y_{i1}\perp Y_{i2}$ است، به این معنا که Y_{i1} و Y_{i2} از هم مستقل‌اند. همچنین ϕ_{i12} در قید زیر صدق می‌کند:

$$\max(0,\phi_{i1}+\phi_{i2}-1)\le\phi_{i12}\le\min(\phi_{i1},\phi_{i2}).$$

متقابلاً بر اساس (۹.۶)، بازه‌هایی برای ω_{i12} و λ_{i12} می‌توانند به دست آیند. نسبت‌های وابسته‌ی مراتب بالاتر نیز به‌صورت زیر تعریف می‌شوند:

$$\omega_{i123} = \frac{\phi_{i123}}{\phi_{i1}\phi_{i2}\phi_{i3}}, \ \ldots, \ \omega_{i123\ldots T} = \frac{\phi_{i123\ldots T}}{\phi_{i1}\phi_{i2}\ldots\phi_{iT}}.$$

نسبت‌های وابسته به‌وسیله‌ی مقایسه‌ی احتمال‌های موفقیت حاشیه‌ای با حالت استقلال تفسیر می‌شوند. اگر نسبت وابسته با دو اندیس و بیش‌تر، برابر با یک باشد (مثلاً اگر $\omega_{i123} = 1$ باشد)، $\phi_{i123} = \phi_{i1}\phi_{i2}\phi_{i3}$ استقلال سه‌تایی Y_{i1}, Y_{i2}, Y_{i3} را نتیجه می‌دهد. بازه‌هایی برای نسبت‌های مراتب بالاتر (مثلاً ϕ_{i123}) با در نظر گرفتن نامساوی $\phi_{i123} \leq \phi_{i12}$ به دست می‌آیند. باید توجه داشت که اگر موفقیت و شکست برای مؤلفه‌های دودویی (Y_{i1} و $1 - Y_{i2}$) مورد مطالعه قرار گیرد، ضریب همبستگی و لگاریتم نسبت بخت در علامت تغییر می‌کنند، اما در مقدار قدر مطلق تغییر نمی‌کنند؛ زیرا

$$\operatorname{corr}(Y_{i1}, 1 - Y_{i2}) = -\operatorname{corr}(Y_{i1}, Y_{i2})$$

$$\begin{aligned}
\operatorname{OR}(Y_{i1}, 1 - Y_{i2}) &= \frac{\Pr(Y_{i1}=1, 1-Y_{i2}=1)\Pr(Y_{i1}=0, 1-Y_{i2}=0)}{\Pr(Y_{i1}=0, 1-Y_{i2}=1)\Pr(Y_{i1}=1, 1-Y_{i2}=0)} \\
&= \frac{\Pr(Y_{i1}=0, Y_{i2}=1)\Pr(Y_{i1}=1, Y_{i2}=0)}{\Pr(Y_{i1}=1, Y_{i2}=1)\Pr(Y_{i1}=0, Y_{i2}=0)} \\
&= \frac{1}{\operatorname{OR}(Y_{i1}, Y_{i2})}.
\end{aligned}$$

در نتیجه:

$$\log[\operatorname{OR}(Y_{i1}, 1 - Y_{i2})] = -\log[\operatorname{OR}(Y_{i1}, Y_{i2})].$$

اما ω_{i12} به $\omega^*_{i12} = \frac{1-\phi_{i2}\omega_{i12}}{1-\phi_{i2}}$ تبدیل می‌شود؛ زیرا

$$\begin{aligned}
\omega^*_{i12} &= \frac{\phi^*_{i12}}{\phi_{i1}\phi_{i2*}} = \frac{E\{Y_{i1}(1 - Y_{i2})\}}{\phi_{i1}(1 - \phi_{i2})} \\
&= \frac{\phi_{i1} - \phi_{i12}}{\phi_{i1} - \phi_{i1}\phi_{i2}} = \frac{1 - \frac{\phi_{i12}}{\phi_{i1}}}{1 - \phi_{i2}} \\
&= \frac{1 - \phi_{i2}\omega_{i12}}{1 - \phi_{i2}}.
\end{aligned}$$

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

از نسبت وابسته برای (Y_{i1}, Y_{i2}) است. بنا بر این، نسبت وابسته که برای همبستگی به کار می‌رود، وابسته به این است که ۱ دلالت بر پیروزی یا شکست در پاسخ‌ها دارد.
حال که شیوه‌ی در نظر گرفتن همبستگی‌ها تشریح شده است، در مثال زیر، طریقه‌ی نوشتن تابع درست‌نمایی بحث می‌شود.

مثال: مدل حاشیه‌ای با لگاریتم نسبت وابسته به‌عنوان پارامتر همبستگی با دو پاسخ برای هر آزمودنی

در این مثال، مدل حاشیه‌ای که از لگاریتم نسبت وابسته برای همبستگی بین پاسخ‌ها استفاده می‌کند، برای تحلیل داده‌های طولی دودویی با دو پاسخ برای هر آزمودنی مورد استفاده قرار می‌گیرد.
تابع درست‌نمایی بر اساس صورت نمایی (۱.٦) و پارامترهای در نظر گرفته‌شده در این مدل، ارائه می‌شود. در صورت نمایی (۱.٦)، $\Psi_i = (\text{logit}(\phi_{i1}), \text{logit}(\phi_{i2}))'$ و $\Omega_i = \log(\omega_{i12}) = \lambda_{i12}$ است؛ بنا بر این صورت نمایی رابطه‌ی (۱.٦) برای iامین فرد، به‌صورت زیر خواهد بود:

$$\Pr(Y_{i1} = y_1, Y_{i2} = y_2) = \exp\{y_1 \text{logit}(\phi_{i1}) + y_2 \text{logit}(\phi_{i2}) + y_1 y_2 \lambda_{i12} - A(\Psi_i, \lambda_{i12})\},$$

که اگر فرض کنیم $\lambda_{i12} = \alpha$، بر اساس روابط زیر می‌توان $A(\Psi_i, \alpha)$ را به دست آورد.

$$\Pr(Y_{i1} = 0, Y_{i2} = 0) = \exp\{-A(\Psi_i, \alpha)\}$$

$$\Pr(Y_{i1} = 1, Y_{i2} = 0) = \exp\{\text{logit}(\phi_{i1}) - A(\Psi_i, \alpha)\}$$

$$\Pr(Y_{i1} = 0, Y_{i2} = 1) = \exp\{\text{logit}(\phi_{i2}) - A(\Psi_i, \alpha)\}$$

$$\Pr(Y_{i1} = 1, Y_{i2} = 1) = \exp\{\text{logit}(\phi_{i1}) + \text{logit}(\phi_{i2}) + \alpha - A(\Psi_i, \alpha)\}$$

و با توجه به این‌که جمع چهار رابطه‌ی بالا برابر با یک می‌باشد، خواهیم داشت:

$$\exp\{-A(\Psi, \alpha)\} = \left(1 + \exp\{\text{logit}(\phi_{i1})\} + \exp\{\text{logit}(\phi_{i2})\} + \exp\{\text{logit}(\phi_{i1}) + \text{logit}(\phi_{i2}) + \alpha\}\right)^{-1}.$$

در نتیجه تابع درست‌نمایی به‌صورت زیر خواهد بود:

$$L = \prod_{i=1}^{n} \Big[\exp\{y_{i1} \text{logit}(\phi_{i1}) + y_{i2} \text{logit}(\phi_{i2}) + y_{i1} y_{i2} \alpha\}$$

$$\times\left(1+\exp\{\mathrm{logit}(\phi_{i1})\}+\exp\{\mathrm{logit}(\phi_{i2})\}+\exp\{\mathrm{logit}(\phi_{i1})+\mathrm{logit}(\phi_{i2})+\alpha\}\right)^{-1}\Bigg]$$

و لگاریتم درست‌نمایی به‌صورت زیر است:

$$\ell=\sum_{i=1}^{n}\Bigg[\{y_{i1}\mathrm{logit}(\phi_{i1})+y_{i2}\mathrm{logit}(\phi_{i2})+y_{i1}y_{i2}\alpha\}$$
$$-\log\left(1+\exp\{\mathrm{logit}(\phi_{i1})\}+\exp\{\mathrm{logit}(\phi_{i2})\}+\exp\{\mathrm{logit}(\phi_{i1})+\mathrm{logit}(\phi_{i2})+\alpha\}\right)\Bigg].$$

لگاریتم درست‌نمایی به مقدار پارامتر β وابسته است که در رابطه‌ی زیر صدق می‌کند:

$$\phi_{it}=\frac{\exp(x'_{it}\beta)}{1+\exp(x'_{it}\beta)}$$

و براوردهای ماکسیمم درست‌نمایی از روش‌های تکراری به دست می‌آیند.

۳.۳.۴.۶ مدل حاشیه‌ای با لگاریتم نسبت بخت حاشیه‌ای به‌عنوان پارامتر همبستگی

این مدل‌بندی توسط لیانگ و دیگران (۱۹۹۲) و لیپ‌شیتس و دیگران (۱۹۹۱) ارائه شده است. در این مدل‌بندی، در معادله‌ی (۱.۶) پارامترهای Ψ_i عبارت‌هایی از احتمال‌های حاشیه‌ای به‌صورت زیر می‌باشند:

$$\psi_{it}=\mathrm{logit}\{\Pr(Y_{it}=1)\},$$

و پارامترهای دوطرفه‌ی $\Omega_i=(\omega_{irs})$ لگاریتم نسبت بخت حاشیه‌ای‌اند؛ یعنی:

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

و پارامترهای سه‌طرفه و بالاتر به‌صورت توابعی از لگاریتم نسبت بخت حاشیه‌ای می‌باشند. به‌عنوان مثال،

$$\exp(\omega_{irsl})=\frac{\mathrm{OR}(Y_{ir},Y_{is}|Y_{il}=1)}{\mathrm{OR}(Y_{ir},Y_{is}|Y_{il}=0)},$$

یا به‌طور معادل،

$$\omega_{irsl}=\log[\mathrm{OR}(Y_{ir},Y_{is}|Y_{il}=1)]-\log[\mathrm{OR}(Y_{ir},Y_{is}|Y_{il}=0)].$$

می‌توان جمله‌ی عمومی پارامترهای ω_i را به‌صورت زیر نوشت که تا حدودی پیچیده است:

$$\omega_{ij_1 j_2 \dots j_s} = \sum(-1)^{(\sum_{l=3}^{s} Y_{il}+s-2)} \log[\mathrm{OR}(Y_{j1}, Y_{j2}|Y_{j3}, \dots, Y_{js})],$$

که جمع‌بندی بر روی همه‌ی 2^{s-2} ترکیب ممکن $(Y_{j3}, \dots, Y_{js})$ می‌باشد.

تفاوت اساسی بین این مدل و مدل لگاریتم نسبت بخت شرطی، در تفسیر دو گشتاور اول آن‌ها است. در لگاریتم نسبت بخت حاشیه‌ای، از احتمال‌های حاشیه‌ای استفاده می‌شود، در حالی که در مدل لگاریتم نسبت بخت شرطی، از احتمال‌های شرطی استفاده می‌شود.

مزیت مدل لگاریتم نسبت بخت حاشیه‌ای نسبت به مدل لگاریتم نسبت بخت شرطی این است که مدل اول (نسبت بخت حاشیه‌ای) قابل تولید شدن دوباره است؛ یعنی اگر $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT})'$ در یک مدل حاشیه‌ای صدق کند، زیرمجموعه‌ای از Y_iها مانند $Y_i' = (Y_{i1}, \dots, Y_{iT'})'$، $(T' < T)$، نیز در مدلی با پارامترهای متناظر آن صدق می‌کند. در نتیجه تفسیر پارامترها در این مدل، صرف نظر از اندازه‌ی بردار Y (یعنی T) و وجود یا عدم پارامترهای همبستگی مراتب بالاتر در مدل، مشابه است. این در حالی است که مدل دومی (مدل نسبت بخت شرطی) قابل تبدیل نیست، زیرا پارامترهای مدل با نسبت بخت حاشیه‌ای به تمام فضای پارامترها مقید نیست، در حالی که پارامترهای مدل با نسبت بخت شرطی بر روی کل پارامترها شرط دارد. همچنین می‌توان گفت که مدل اول برای پاسخ‌های تکرار شده‌ی با اندازه‌ی متفاوت، قابل استفاده است، در حالی که به کار بردن مدل دوم برای چنین داده‌هایی امکان‌پذیر نیست.

همان‌طور که ملاحظه می‌شود، این گونه مدل‌ها نیاز به محاسبات زیادی برای براورد کردن پارامترها دارند (به‌جز برای $T = 2$). در نتیجه برای رهایی از این مشکل، محققان اکثراً همبستگی‌های مراتب بالاتر را صفر در نظر می‌گیرند. به‌عنوان مثال، ژائو و پرنتیس (۱۹۹۰) مدلی با صورت نمایی درجه‌ی دوم با پارامترهای Ψ_i و $\Omega_i = (\omega_{i12}, \dots, \omega_{iT-1T})'$ را در نظر گرفته‌اند. بر اساس این مدل‌بندی، فقط همبستگی مرتبه‌ی دوم در نظر گرفته می‌شود. به‌عنوان مثالی دیگر، اگر Ω_i بردار صفر باشد، مدل استقلال به دست خواهد آمد.

مثال: مدل حاشیه‌ای با پارامتر همبستگی به‌صورت لگاریتم نسبت بخت حاشیه‌ای با دو پاسخ برای هر آزمودنی

در این مثال، از مدل حاشیه‌ای استفاده شده است که با در نظر گرفتن لگاریتم نسبت بخت حاشیه‌ای برای همبستگی بین پاسخ‌ها، به تحلیل داده‌های طولی با دو پاسخ برای هر آزمودنی پرداخته شده است.

بر اساس معادله‌ی (۱.۶) و پارامترهای در نظر گرفته‌شده بر اساس این روش داریم:

$$f(y_i) = \Pr(Y_{i1} = y_{i1}, Y_{i2} = y_{i2})$$
$$= \exp\{y_{i1}\psi_{i1} + y_{i2}\psi_{i2} + y_{i1}y_{i2}\omega_{i12} - A\{\Psi_i, \omega_{i12}\}\},$$

که در آن ω_{i12} و ψ_{it} $(t = 1, 2)$ به‌صورت زیر تعریف می‌شوند:

$$\psi_{it} = logit\{\Pr(Y_{it} = 1)\},$$
$$\exp(\omega_{i12}) = \frac{\Pr(Y_{i1} = 1, Y_{i2} = 1)\Pr(Y_{i1} = 0, Y_{i2} = 0)}{\Pr(Y_{i1} = 1, Y_{i2} = 0)\Pr(Y_{i1} = 0, Y_{i2} = 1)}.$$

اگر فرض کنیم $\omega_{i12} = \alpha$، تابع درست‌نمایی به‌صورت زیر در می‌آید:

$$L = \prod_{i=1}^{n} \exp\{y_{i1}\psi_{i1} + y_{i2}\psi_{i2} + y_{i1}y_{i2}\alpha - A(\Psi_i, \alpha)).$$

با استفاده از این اطلاع که مجموع احتمال‌های مشترک همه‌ی حالت‌های Y_i، یک است و همچنین با استفاده از روابط زیر می‌توان مقدار $A(\Psi_i, \alpha)$ را به دست آورد.

$$\Pr(Y_{i1} = 0, Y_{i2} = 0) = \exp\{-A(\Psi_i, \alpha)\},$$

$$\Pr(Y_{i1} = 1, Y_{i2} = 0) = \exp\{\psi_{i1} - A(\Psi_i, \alpha)\},$$

$$\Pr(Y_{i1} = 0, Y_{i2} = 1) = \exp\{\psi_{i2} - A(\Psi_i, \alpha)\},$$

$$\Pr(Y_{i1} = 1, Y_{i2} = 1) = \exp\{\psi_{i1} + \psi_{i2} + \alpha - A(\Psi_i, \alpha)\}.$$

بنا بر این با توجه به یک بودن جمع چهار رابطه‌ی بالا داریم

$$\exp\{-A(\Psi 0, \alpha)\} = \Big(1 + \exp(\psi_{i1}) + \exp(\psi_{i2}) + \exp(\psi_{i1} + \psi_{i2} + \alpha)\Big)^{-1}.$$

در نتیجه تابع درست‌نمایی به‌صورت زیر در می‌آید:

$$L = \prod_{i=1}^{n} \Big[\exp\{y_{i1}\psi_{i1} + y_{i2}\psi_{i2} + y_{i1}y_{i2}\alpha\}$$

$$\times\Big(1+\exp(\psi_{i1})+\exp(\psi_{i2})+\exp(\psi_{i1}+\psi_{i2}+\alpha)\Big)^{-1}$$

و لگاریتم درست‌نمایی به‌صورت زیر است:

$$\ell=\sum_{i=1}^{n}\Big[\Big(y_{i1}\psi_{i1}+y_{i2}\psi_{i2}+y_{i1}y_{i2}\alpha\Big) - \log\Big(1+\exp(\psi_{i1})+\exp(\psi_{i2})+\exp(\psi_{i1}+\psi_{i2}+\alpha)\Big)\Big],$$

که این تابع لگاریتم درست‌نمایی، همانند لگاریتم درست‌نمایی مثال نسبت وابسته، به پارامترهای β وابسته است.

۴.۳.۴.۶ مدل‌های حاشیه‌ای با استفاده از مفهوم متغیر پنهان

با استفاده از مفهوم متغیر پنهان که در فصل‌های قبل بیان شد، مدل‌های حاشیه‌ای برای تحلیل داده‌های طولی دودویی شرح داده می‌شوند. در این گونه مدل‌ها چون Y^*_{it} متغیری پیوسته است، دسته‌ای از مدل‌های خطی را می‌توان برای تحلیل آن‌ها در نظر گرفت، که به‌صورت زیر تعریف می‌شوند:

$$Y^*_{it}=x'_{it}\beta_t+\varepsilon_{it},$$

که در آن فرض‌هایی در باره‌ی ε_{it}ها در نظر گرفته می‌شود (به فصل اول رجوع شود). به‌عنوان مثال، $\varepsilon_i=(\varepsilon_{i1},\dots,\varepsilon_{it})'$ برای آزمودنی iام، نرمال T—متغیره با میانگین صفر و ماتریس کوواریانس

$$\Sigma=\begin{pmatrix}1&\rho_{12}&\cdots&\rho_{1T}\\ \rho_{21}&1&\cdots&\rho_{2T}\\ \vdots&\vdots&&\vdots\\ \rho_{T1}&\rho_{T2}&\cdots&1\end{pmatrix}=R$$

در نظر گرفته می‌شود.

با توجه به این‌که Y_{it}های مشاهده‌شده، بر اساس قاعده‌ی زیر مشخص می‌شوند

$$y_{it}=\begin{cases}1, & y^*_{it}>a\\ 0, & y^*_{it}\le a\end{cases}$$

که $a = \circ$ اختیار می‌شود (به فصل اول رجوع شود)، احتمال‌های مشترک و در نتیجه تابع درست‌نمایی با استفاده از فرض‌های روی ε_i قابل محاسبه‌اند. به‌عنوان مثال، برای iامین فرد، احتمال موفقیت مشترک پاسخ‌ها به‌صورت زیر خواهد بود:

$$\begin{aligned}
P_{i1\cdots 1} &= \Pr(Y_{i1} = 1, \ldots, Y_{iT} = 1) = \Pr(Y^*_{i1} > 0, \ldots, Y^*_{iT} > 0) \\
&= \Pr(x'_{i1}\beta_1 + \varepsilon_{i1} > 0, \ldots, x'_{iT}\beta_T + \varepsilon_{iT} > 0) \\
&= \Pr(\varepsilon_{i1} > -x'_{i1}\beta_1, \ldots, \varepsilon_{iT} > -x'_{iT}\beta_T) \\
&= \Pr(\varepsilon_{i1} \le x'_{i1}\beta_1, \ldots, \varepsilon_{iT} \le x'_{iT}\beta_T) \\
&= \int_{-\infty}^{x'_{i1}\beta_1} \cdots \int_{-\infty}^{x'_{iT}\beta_T} f_{\varepsilon_{i1},\ldots,\varepsilon_{iT}}(z_{i1}, \ldots, z_{iT}) dz_{i1} \ldots dz_{iT} \\
&= \Phi_T(x'_{i1}\beta_1, \ldots, x'_{iT}\beta_T; R),
\end{aligned}$$

که در آن $f_{\varepsilon_{i1},\ldots,\varepsilon_{iT}}(.)$ تابع چگالی مشترک نرمال T–متغیره و Φ_T تابع توزیع تجمعی آن است. به همین صورت، $P_{ijk\cdots l}$ به‌ازای هر $j, k, \ldots, l = 0, 1$ نیز به دست می‌آید. در نتیجه تابع درست‌نمایی، حاصل‌ضرب چنین احتمالاتی برای همه‌ی آزمودنی‌ها است. همان‌طور که ملاحظه می‌شود با افزایش T، تعداد پارامترهای ماتریس همبستگی افزایش چشم‌گیری خواهد داشت، $\frac{T(T-1)}{2}$، که براورد همه‌ی این پارامترها نیاز به محاسبات و وقت زیاد دارد. به همین دلیل برای ماتریس همبستگی، حالت‌های خاصی در نظر گرفته می‌شود. به‌عنوان مثال،

۱) $\mathrm{corr}(\varepsilon_{it}, \varepsilon_{it'}) = \rho, \forall t \neq t'$

۲) $\mathrm{corr}(\varepsilon_{it}, \varepsilon_{it'}) = \alpha^{|t-t'|}$

در نظر گرفتن این فرض‌ها برای ماتریس همبستگی، تعداد پارامترهای موجود در R را کاهش می‌دهد.

مثال: مدل حاشیه‌ای با استفاده از مفهوم متغیر پنهان برای داده‌های طولی با دو پاسخ برای هر آزمودنی

با توجه به صورت کلّی این گونه مدل‌ها و با در نظر گرفتن $T = 2$، خطاهای ε_iها دارای توزیع نرمال دومتغیره با میانگین صفر و ماتریس کوواریانس زیرند.

$$\Sigma_i = \mathrm{var}(\varepsilon_i) = \begin{bmatrix} 1 & \mathrm{corr}(Y_{i1}, Y_{i2}) \\ \mathrm{corr}(Y_{i1}, Y_{i2}) & 1 \end{bmatrix}$$

اگر فرض شود $\rho = \mathrm{corr}(\varepsilon_{i1}, \varepsilon_{i2})$، احتمال موفقیت مشترک دو پاسخ برای هر آزمودنی به‌صورت زیر به دست می‌آید:

$$
\begin{aligned}
P_{i11} &= \Pr(Y_{i1} = 1, Y_{i2} = 1) \\
&= \Pr(Y_{i1}^{*} > 0, Y_{i2}^{*} > 0) \\
&= \Pr(x_{i1}'\beta_1 + \varepsilon_{i1} > 0, x_{i2}'\beta_2 + \varepsilon_{i2} > 0) \\
&= \Pr(\varepsilon_{i1} > -x_{i1}'\beta_1, \varepsilon_{i2} > -x_{i2}'\beta_2) \\
&= \Pr(\varepsilon_{i1} \leq x_{i1}'\beta_1, \varepsilon_{i2} \leq x_{i2}'\beta_2) \\
&= \int_{-\infty}^{x_{i1}'\beta_1} \int_{-\infty}^{x_{i2}'\beta_2} f_{\varepsilon_{i1},\varepsilon_{i2}}(z_{i1}, z_{i2}) dz_{i1} dz_{i2} \\
&= \phi_2(x_{i1}'\beta_1, x_{i2}'\beta_2; \rho),
\end{aligned}
$$

که در آن $f_{\varepsilon_{i1},\varepsilon_{i2}}(.)$ تابع چگالی مشترک نرمال دومتغیره و Φ_2 تابع توزیع تجمعی $(\varepsilon_{i1}, \varepsilon_{i2})$ است. به همین ترتیب P_{i00} و P_{i10} و P_{i01} نیز محاسبه می‌شوند. در نتیجه، تابع درست‌نمایی به‌صورت زیر است:

$$\prod_{i=1}^{n} P_{i11}^{y_{i1}y_{i2}} P_{i00}^{(1-y_{i1})(1-y_{i2})} P_{i01}^{(1-y_{i1})y_{i2}} P_{i10}^{y_{i1}(1-y_{i2})}. \qquad (۱۰.۶)$$

روش دوم برای در نظر گرفتن ساختار همبستگی، استفاده از مدل‌بندی ساختار همبستگی به‌جای استفاده از نسبت بخت‌ها است. در این روش، مؤلفه‌ی سیستماتیک به‌وسیله‌ی دو متغیر پنهان Y_{i1}^{*} و Y_{i2}^{*} که از توزیع نرمال دومتغیره آمده‌اند، به‌صورت زیر تعریف می‌شوند. احتمال‌های توأم دو پاسخ می‌توانند از طریق توزیع توأم Y_{it}^{*}ها به‌صورت زیر تعیین شوند:

$$\Pr(Y_{i1} = 1, Y_{i2} = 1) = \Pr(Y_{i1}^{*} > 0, Y_{i2}^{*} > 0) = \int_{0}^{\infty} \int_{0}^{\infty} f(Y_i^{*} | \beta, \Sigma_i),$$

که در آن

$$f(Y_i^{*} | \beta, \Sigma_i) = \frac{1}{(2\pi)|\Sigma_i|^{\frac{1}{2}}} \exp\{\frac{-1}{2}(Y_i^{*} - \mu_i(\beta))^t \Sigma_i^{-1} (Y_i^{*} - \mu_i(\beta))\}.$$

در حالت دومتغیره با اِعمال یک تبدیل یک‌به‌یک روی ρ که توسط لوسزی و هاولینگن (۱۹۹۴) معرفی شده است، داریم:

$$\log(\frac{1 + \rho_i}{1 - \rho_i}) = \alpha_0 + \alpha_1 X_i,$$

که در آن $\rho_i = \mathrm{cov}(Y_{i1}, Y_{i2})$ است و X_i متغیر کمکی مناسب است.

براوردهای ماکسیمم درست‌نمایی پارامترهای مدل‌های ذکر شده در این زیربخش بر اساس روش‌های تکراری عددی به دست می‌آیند. برای همه‌ی مدل‌های حاشیه‌ای که در این زیربخش بیان شده‌اند، یعنی مدل‌هایی که پارامترهای همبستگی را به‌صورت توابعی از لگاریتم نسبت بخت شرطی یا لگاریتم نسبت بخت حاشیه‌ای یا نسبت وابسته در نظر گرفته‌اند، بدون توجه به ساختار همبستگی، براورد پارامترهای رگرسیونی، سازگار می‌باشند؛ ولی اگر ساختار همبستگی برای این مدل‌ها به‌درستی مشخص شده باشند، براورد $\mathrm{cov}(\hat{\beta})$ نیز براوردی سازگار است.

رهیافت GEE و بسط آن که با GEE2 نشان داده شده است، رهیافت‌هایی هستند که بر اساس فرض‌هایی بر مبنای گشتاورهای اول و دوم پاسخ‌ها به دست می‌آیند و نیاز به مشخص کردن توزیع کامل پاسخ‌ها ندارند. باید توجه داشت که در توزیع گاوسی، فرض در باره‌ی دو گشتاور اول، توزیع را به‌طور کامل مشخص می‌کند؛ ولی برای داده‌های دودویی که مورد بحث اصلی است، این مطلب برقرار نمی‌باشد. بر اساس قضیه‌ی ۱.۶، $\hat{\beta}_G$ براوردگری سازگار برای β است و $n^{\frac{1}{2}}(\hat{\beta}_G - \beta)$ به‌طور مجانبی، گاوسی با میانگین صفر و ماتریس کوواریانس مشخص‌شده در این قضیه می‌باشد. تفاوت اساسی بین GEE و GEE2 این است که براوردگر $\hat{\beta}_G$ صرف نظر از ساختار همبستگي در نظر گرفته شده بین پاسخ‌ها، سازگار است؛ اما اگر ساختار همبستگی به‌درستی مشخص‌نشده باشد، براورد β تحت GEE2 ناسازگار خواهد بود. ولی اگر ساختار همبستگی به‌درستی مشخص شده باشد، براوردگر β تحت GEE2 علاوه بر سازگاری، دارای کارایی بیش‌تری نسبت به $\hat{\beta}_G$ خواهد بود (لیانگ و زیگر، ۱۹۸۶).

۵.۳.۴.۶ مدل اثرهای تصادفی

طریقه‌ی محاسبه‌ی تابع درست‌نمایی در مدل اثرهای تصادفی برای آزمودنی iاُم به‌صورت زیر است:

$$f(y_{i1}, \ldots, y_{iT_i}) = \int \prod_{t=1}^{T_i} f(y_{it}|\tau_i) f(\tau_i) d(\tau_i),$$

که در آن فرض شده است برای آزمودنی iام، T_i پاسخ مشاهده شده و وابستگی به x_i برای سادگی در نوشتار حذف شده است. لگاریتم درست‌نمایی برای n آزمودنی به‌صورت زیر است:

$$\ell(\beta,\phi)=\sum_{i=1}^{n}\log\left\{\int\prod_{t=1}^{T_i}f(y_{it}|\tau_i)g(\tau_i)d(\tau_i)\right\}. \qquad (11.6)$$

برای محاسبه‌ی انتگرال فوق باید از فرمول‌های مربع‌بندی گاوسی استفاده کنیم، به این صورت که تعدادی نقاط جرم (مثلاً k تا) با احتمال‌های متناظر با آن را در نظر می‌گیریم. با توجه به معلوم بودن یا نبودن مقادیر k و احتمال‌های متناظر با آن‌ها، با سه روش برای محاسبه‌ی تابع درست‌نمایی مواجه هستیم:

آ) روش پارامتری: در این روش، مقدار k و احتمال‌های متناظر با آن معلوم فرض می‌شوند.

ب) روش ناپارامتری: در این روش، مقدار k معلوم و احتمال‌های متناظر با آن نامعلوم فرض می‌شوند.

پ) روش کاملاً ناپارامتری: در این روش، مقدار k و احتمال‌های متناظر با آن نامعلوم فرض می‌شوند.

در روش‌های ناپارامتری و کاملاً ناپارامتری مقادیر نامعلوم به‌عنوان پارامتر در نظر گرفته می‌شوند و انتگرال رابطه‌ی (۱۱.۶) به‌صورت زیر تبدیل می‌شود:

$$\ell(\beta,\phi)=\sum_{i=1}^{n}\log\left[\sum_{k}\left\{\prod_{t=1}^{T_i}f(y_{it}|\tau_i=\tau_k)\right\}\pi_k\right], \qquad (12.6)$$

به‌طوری که π_k می‌تواند به‌عنوان احتمال‌های معلوم (حالت «آ») مورد استفاده قرار گیرد یا به‌عنوان پارامتر (حالت‌های «ب» و «پ») براورد شود.

۶.۳.۴.۶ مدل‌های انتقالی

معادلات درست‌نمایی برای مدل‌های انتقالی به‌صورت زیر است:

$$f(y_{i1},\dots,y_{iT_i})=f(y_{i1})f(y_{i2}|y_{i1})f(y_{i3}|y_{i1},y_{i2})\cdots f(y_{iT_i}|y_{i1},\dots,y_{iT_i-1}),$$

که در آن‌ها برای یک مدل انتقالی از مرتبه‌ی q داریم:

$$f(y_{it}|y_{t1},\dots,y_{i,t-q})=(p_{it})^{y_{it}}(1-p_{it})^{1-y_{it}},\qquad t=1,\dots,T_i,\qquad i=1,\dots,n$$

که در آن

$$p_{it} = \left[\frac{\exp(x'_{it}\beta_t + \sum_{j=۱}^{q} \gamma_{jt} y_{it-j})}{۱ + \exp(x'_{it}\beta_t + \sum_{j=۱}^{q} \gamma_{jt} y_{it-j})}\right].$$

۵.۶ مثال‌های کاربردی

در این بخش، سه مثال کاربردی مورد بحث و تحلیل قرار می‌گیرند.

۱.۵.۶ داده‌های مربوط به آسم

در این داده‌ها برازش مدل حاشیه‌ای با استفاده از مفهوم متغیر پنهان صورت می‌گیرد. برای بررسی تأثیر سن و جنس بر متغیر پاسخ که ابتلا یا عدم ابتلا به بیماری آسم است، مدل زیر در نظر گرفته شده است:

$$Y^*_{it} = \alpha + \beta_۱ \text{sex} + \beta_۲ I(\text{age}۱۳) + \varepsilon_{it}, \qquad t = ۱, ۲$$

در این مدل، $I(\text{age}۱۳)$ یک تابع نشانگر است که به‌صورت زیر تعریف می‌شود:

$$I(age۱۳) = \begin{cases} ۱, & \text{اگر فرد سیزده ساله باشد} \\ ۰, & \text{اگر فرد سیزده ساله نباشد} \end{cases}$$

همچنین متغیرهای پنهان Y_{it}، برای $Y^*_{i۱}$ و $Y^*_{i۲}$، به‌صورت زیر تولید می‌شوند:

$$Y_{it} = \begin{cases} ۱, & Y^*_{it} > ۰ \\ ۰, & Y^*_{it} \leq ۰ \end{cases}$$

در سن ۱۳ سالگی، $I(\text{age}۱۳)$ وارد مدل می‌شود که پارامتر مربوط به آن نشان‌دهنده‌ی تأثیر زمان بر ابتلا یا عدم ابتلا به آسم است. خطاهای $(\varepsilon_{i۱}, \varepsilon_{i۲})'$ دارای توزیع نرمال دومتغیره با میانگین صفر، واریانس یک و همبستگی ρ می‌باشند.
در این مثال، یک بار دو متغیر $Y_{i۱}$ و $Y_{i۲}$ مستقل فرض شده‌اند و بار دیگر همبستگی بین

دو متغیر در نظر گرفته شده است. تابع درست‌نمایی برای حالت کلّی به‌صورت زیر است:

$$L = \prod_{i=1}^{n} P_{i11}^{y_{i1}y_{i2}} P_{i00}^{(1-y_{i1})(1-y_{i2})} P_{i01}^{(1-y_{i1})y_{i2}} P_{i10}^{y_{i1}(1-y_{i2})},$$

که درآن،

$$P_{i00} = \Pr(Y_{i1} = 0, Y_{i2} = 0) = \Phi_2(-\alpha - \beta_1 \text{sex}, -\alpha - \beta_1 \text{sex} - \beta_2; \rho),$$

$$P_{i01} = \Pr(Y_{i1} = 0, Y_{i2} = 1) = \Phi_2(-\alpha - \beta_1 \text{sex}, \alpha + \beta_1 \text{sex} + \beta_2; -\rho),$$

$$P_{i10} = \Pr(Y_{i1} = 1, Y_{i2} = 0) = \Phi_2(\alpha + \beta_1 \text{sex}, -\alpha - \beta_1 \text{sex} - \beta_2; -\rho),$$

$$P_{i11} = \Pr(Y_{i1} = 1, Y_{i2} = 1) = \Phi_2(\alpha + \beta_1 \text{sex}, \alpha + \beta_1 \text{sex} + \beta_2; \rho),$$

که برای مدل استقلال، $\rho = 0$ در نظر گرفته می‌شود.

با استفاده از روش‌های براورد، براوردهای ماکسیمم درست‌نمایی پارامترها محاسبه و نتایج آن در جدول ۱.٦ داده شده است.

همان‌طور که براورد پارامترها در مدل کامل (عدم استقلال) نشان می‌دهد، داده‌ها شواهد کافی مبنی بر تأثیر زمان در ابتلا به آسم دارند ($P = ۰/۰۰۱$–مقدار). همچنین در این مدل، داده‌ها نشان می‌دهند که جنس در ابتلا به آسم تأثیر دارد، ولی این تأثیر، چندان قوی نیست ($P = ۰/۰۴۳$–مقدار).

جدول ۱.٦: براورد پارامترها تحت مدل‌های در نظر گرفته‌شده

پارامتر	استقلال		عدم استقلال	
	براورد	P–مقدار	براورد	P–مقدار
α	۱/۹	–	۱/۸۹–	–
β_1 (جنس)	۰/۲۴	۰/۰۰۹	۰/۲۳	۰/۰۴۳
β_2 (سن ۱۳ سالگی)	۰/۱۷	۰/۰۵۷	۰/۱۷	۰/۰۰۱
ρ	۰	–	۰/۹۳	۰

مقایسه بین مدل استقلال و عدم استقلال، حاکی از اهمیت براورد پارامتر ρ می‌باشد. همان‌طور که ملاحظه می‌شود، هرچند مقدار براورد پارامترها در دو مدل، نزدیک به یکدیگر است، حضور پارامتر ρ در مدل، روی خطاهای استاندارد براوردگرها تأثیر گذاشته است. خطای استاندارد پارامتر متغیر کمکی سن (زمان مانا) میل به کم براورد شدن دارد (که کم براورد شدن P–مقدار را نتیجه می‌دهد) و خطای استاندارد پارامتر مربوط به متغیر جنس، میل به بیش براورد شدن دارد (که P–مقدار را بیش‌تر نشان می‌دهد).

۲.۵.۶ اثر آلودگی هوا بر سلامت کودکان

در این مثال، از داده‌های مربوط به بررسی اثر آلودگی هوا بر سلامت کودکان استفاده می‌کنیم که توسط ویر و دیگران (۱۹۸۴) گردآوری شده است و در فصل اول به معرفی آن پرداخته بودیم. در این مثال، براوردهای ماکسیمم درست‌نمایی پارامترها برای مدل‌های دارای لگاریتم نسبت بخت شرطی و استقلال آورده شده است و خواهیم دید که در نظر نگرفتن همبستگی، چه تأثیری بر استنباط پارامترها خواهد داشت. در این مثال، اثر سیگار کشیدن مادر بر روی سلامت کودک‌اش مورد بررسی قرار گرفته است. اگرچه سیگاری بودن مادر، متغیر کمکی‌ای است که با زمان تغییر می‌کند، در این‌جا به‌صورت مقداری ثابت در اولین سال بررسی در نظر گرفته شده است.

مدل حاشیه‌ای زیر برای تحلیل این داده‌ها در نظر گرفته شده است:

$$\text{logit}(\mu_{it}) = \beta_۰ + \beta_۱\text{age} + \beta_۲\text{sm} + \beta_۳(\text{age} \times \text{sm})$$

که sm نشان‌دهنده‌ی سیگاری بودن مادر و $\text{age} \times \text{sm}$ اثر متقابل سن و سیگاری بودن مادر است. sm مقادیر صفر و یک را به‌صورت زیر اختیار می‌کند:

$$sm = \begin{cases} ۱, & \text{اگر مادر سیگاری باشد} \\ ۰, & \text{اگر مادر سیگاری نباشد} \end{cases}$$

فیتس‌موریس و لرد (۱۹۹۳) همبستگی را با استفاده از توابع لگاریتم نسبت بخت شرطی مدل‌بندی کردند؛ به این معنا که بردار همبستگی، یعنی $\Omega_i = (\omega_{i۱۲}, \ldots, \omega_{i۳۴}, \ldots, \omega_{i۱۲۳۴})'$ را به‌صورت توابعی از لگاریتم نسبت بخت شرطی در نظر می‌گیرند. برای این داده‌ها دو مدل زیر در نظر گرفته شده است:

۱) مدل استقلال، $\Omega_i = ۰$،

۲) مدل همبستگی قابل تعویض، $\Omega = (\alpha, \alpha, \alpha, \alpha, \alpha, ۰, \ldots, ۰)'$.

نتایج استفاده از این مدل‌ها در جدول ۲.۶ نشان می‌دهد که نرخ بیماری برای کودکان با افزایش سن، کاهش می‌یابد و برای کودکان با مادران سیگاری نسبت به مادران غیرسیگاری افزایش می‌یابد. در این جدول، براوردهای β برای دو مدل، بسیار شبیه به هم هستند، با این تفاوت که برای مدل استقلال، خطاهای استاندارد براورد پارامترها برای اثرهای زمان

مانا (sm) میل به کم براورد شدن و برای سن (age) میل به بیش براورد شدن دارند. در این جدول، G^2 مربوط به مدل‌ها نیز محاسبه شده است که حکایت از نامناسب بودن مدل استقلال می‌کند. برای جزئیات بیش‌تر به فیتس‌موریس و لرد (۱۹۹۳) مراجعه کنید.

در این مثال نشان داده شد که اگر همبستگی موجود در میان پاسخ‌ها در نظر گرفته نشود، با آن‌که در براورد پارامترها خللی ایجاد نمی‌شود، از هر گونه استنباط در مورد پارامترها تأثیر می‌پذیرد و ممکن است استنباطی بر خلاف واقعیت برای پارامترها به دست آید.

جدول ۲.۶: براورد پارامترها تحت مدل‌های در نظر گرفته شده

	مدل استقلال		مدل همبستگی قابل تعویض	
پارامتر	براورد	انحراف معیار	براورد	انحراف معیار
β_0	۱/۹۰۱−	۰/۰۸۹	۱/۹۰۱−	۰/۱۱۷
β_1	۰/۱۴۱−	۰/۰۷	۰/۱۴۱−	۰/۰۵۷
β_2	۰/۳۱۴	۰/۱۳۹	۰/۳۱۴	۰/۱۸۸
β_3	۰/۰۷۱	۰/۱۱۱	۰/۰۷۱	۰/۰۸۷
α	—	—	۱/۲۶۶	۰/۰۷۳
	$G^2 =$ ۲۴۱/۰۸	$d.f. =$ ۲۶	$G^2 =$ ۱۶/۳۹	$d.f. =$ ۲۵

همان‌طور که قبلاً بیان شد، در مدل حاشیه‌ای می‌توان همبستگی را به‌طور مجزا از مدل حاشیه‌ای مدل‌بندی کرد. در مثال زیر که در آن، سه روش مدل حاشیه‌ای، اثرهای تصادفی و مدل انتقالی مقایسه شده‌اند، در مدل حاشیه‌ای از این نوع مدل‌بندی استفاده شده است.

۳.۵.۶ داده‌های بی‌خوابی

در این بخش، از مدل‌های مختلف برای تحلیل اندازه‌گیری‌های مکرر با پاسخ‌های دودویی طولی استفاده می‌کنیم. انتخاب مدل، بسته به نوع سؤال‌های علمی مورد علاقه‌ی فرد می‌تواند انتخاب شود. همان‌طور که قبلاً اشاره شد، مدل حاشیه‌ای هنگامی مورد استفاده قرار می‌گیرد که علاقه‌مند به در نظر گرفتن میانگین جامعه باشیم. مدل اثرهای تصادفی، زمانی مورد استفاده قرار می‌گیرد که علاقه‌مند به دانستن اثر فرد یا خوشه در مدل باشیم، و از مدل انتقالی نیز هنگامی استفاده می‌شود که تمایل به دانستن اطلاعاتی در باره‌ی اثر پاسخ قبلی بر میانگین شرطی پاسخ در زمان حال داشته باشیم. بنا بر این، در این بخش به معرفی سه مدل اشاره‌شده با استفاده از داده‌های بی‌خوابی می‌پردازیم.

۱.۳.۵.۶ مدل‌های حاشیه‌ای

در مدل حاشیه‌ای، رگرسیون پاسخ روی متغیر کمکی به‌طور مجزا از همبستگی بین افراد مدل‌بندی می‌شود. فرض کنید که دانستن همبستگی بین بردار پاسخ و متغیرهای کمکی در هر زمان و این‌که این همبستگی چگونه در طول زمان تغییر می‌کند، مورد علاقه‌ی ما باشد. برای هر مشاهده، دو متغیر پاسخ دودویی Y_1 و Y_2 تعریف می‌کنیم که هر یک از آن‌ها مقادیر ۰ $(Y_1 < 30)$ یا ۱ $(Y_1 \geq 30)$ را اخذ می‌کند. در مدل‌های حاشیه‌ای، دو روش برای در نظر گرفتن همبستگی بین پاسخ‌ها وجود دارد. اولین روش این است که پاسخ توأم (Y_1, Y_2) را با استفاده از احتمال‌های حاشیه‌ای برای هر متغیر پاسخ و نسبت بخت که همبستگی بین دو متغیر وابسته را پارامتری می‌کند، مدل‌بندی کنیم. مؤلفه‌های سیستماتیک، نسبت بخت را به همان خوبی مدل‌بندی می‌کنند که احتمال‌های حاشیه‌ای $\pi_{tj} = \Pr(Y_t = j|x_t)$ را. نسبت بخت که به‌صورت $\varphi = \frac{\pi_{(1,2)(0,0)}\pi_{(1,2)(1,1)}}{\pi_{(1,2)(1,0)}\pi_{(1,2)(0,1)}}$ معرفی می‌شود، همبستگی بین دو پاسخ را توصیف می‌کند، که در آن $\pi_{(1,2)(j_1,j_2)} = \Pr(Y_1 = j_1, Y_2 = j_2)$. همچنین برای هر مشاهده به‌ازای $t = 1, 2$ داریم $\Pr(Y_t = 1|x_t) = \pi_{t1} = \frac{\exp(x_t'\beta_t)}{1+\exp(x_t'\beta_t)}$، که در آن x_t بردار متغیرهای کمکی است و برای همبستگی بین دو پاسخ داریم $\log(\varphi) = x_3'\beta_3$، که در آن نیز x_3 بردار متغیر کمکی است و φ ساختار همبستگی بین اندازه‌های مکرر است. این مدل، اولین بار توسط دیل (۱۹۸۶) معرفی شد. همان‌طور که در مثال‌های قبل دیدیم، اگر ساختار همبستگی در نظر گرفته نشود، براورد انحراف معیار براوردگرهای پارامترها ممکن است به‌درستی گزارش نشود (لیانگ و زیگر، ۱۹۸۶، نیز به این مطالب اشاره کرده‌اند). بنا بر این مدل معرفی‌شده برای داده‌های بی‌خوابی به‌صورت زیر است:

$$\text{logit}[\Pr(Y_{it} = 1|\alpha_t, \beta_t, \text{treat}_i)] = \alpha_t + \beta_t \times \text{treat}_i, \qquad t = 1, 2$$

$$\varphi = \exp(\alpha_3 + \beta_3 \times \text{treat}_i),$$

که در آن treat_i متغیر نشانگر برای درمان است (۱: داروی مؤثر، ۰: دارونما). براورد پارامترها به‌روش ماکسیمم درست‌نمایی با استفاده از مدل‌بندی حاشیه‌ای پاسخ‌های دومتغیره با و بدون در نظر گرفتن ساختار همبستگی در جدول ۳.۶ آمده است (در مدل

بدون در نظر گرفتن همبستگی $\alpha_3 = \beta_3 = 0$، مدل استقلال، و در مدل با در نظر گرفتن همبستگی ۱، $\beta_3 = 0$ و α_3 براورد شده است و در مدل با در نظر گرفتن همبستگی ۲، هر دو پارامتر β_3 و α_3 براورد شده‌اند).

جدول ۳.۶: براورد پارامترهای مدل حاشیه‌ای با استفاده از نسبت بخت‌ها با و بدون در نظر گرفتن ساختار همبستگی برای داده‌های بی‌خوابی (همبستگی ۲: همبستگی‌ای که به متغیرهای کمکی وابسته است، همبستگی ۱: همبستگی‌ای که به متغیرهای کمکی وابسته نیست)

	با در نظر گرفتن همبستگی ۲		با در نظر گرفتن همبستگی ۱		بدون در نظر گرفتن همبستگی	
پارامتر	براورد	انحراف معیار	براورد	انحراف معیار	براورد	انحراف معیار
α_1	۰/۹۲۸	۰/۱۷۸	۰/۹۲۸	۰/۲۰۳	۰/۹۲۸	۰/۲۰۳
α_2	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۱۸۳	۰/۰۰۰	۰/۱۸۳
α_3	۲/۶۳۹	۰/۵۷۸	۱/۹۰۰	۰/۴۰۰	—	—
β_1: درمان (داروی مؤثر)	۰/۰۷۲	۰/۲۷۶	۰/۰۷۲	۰/۲۸۹	۰/۰۷۲	۰/۲۸۹
β_2: درمان (داروی مؤثر)	−۱/۰۸۷	۰/۲۱۱	−۱/۰۸۷	۰/۲۸۰	−۱/۰۸۷	۰/۲۸۰
β_3: درمان (داروی مؤثر)	−۱/۸۶۱	۰/۷۹۲	—	—	—	—

نتایج مدل حاشیه‌ای در جدول ۳.۶ برای پاسخ زمان اول نشان می‌دهد که هیچ اثر معناداری از نوع درمان روی احتمال پاسخ زمان اول وجود ندارد ($P = 0/803$–مقدار)، اما اثر معناداری از نوع درمان روی احتمال پاسخ در زمان دوم وجود دارد ($P = 0/000$–مقدار). همچنین ساختار همبستگی بین پاسخ‌ها معنادار است ($P = 0/000$–مقدار). همبستگی بین پاسخ‌ها که از طریق پارامتر α_3 نمایش داده می‌شود، نشان می‌دهد که هرچه مدت خوابیدن در زمان اول بیش‌تر باشد، مدتِ به خواب رفتن در زمان دوم نیز بیش‌تر می‌شود.

روش دوم برای در نظر گرفتن ساختار همبستگی، استفاده از مدل‌بندی ساختار همبستگی به‌جای استفاده از نسبت بخت‌ها است که در فصل‌های قبل به آن پرداخته شد. بنا بر این دو مدل رگرسیونی برای داده‌های بی‌خوابی به‌صورت زیر تعریف می‌شود، که اولی برای احتمال‌های حاشیه‌ای (با در نظر گرفتن تابع ربط پروبیت) است و دومی برای ساختار همبستگی.

$$\Pr(Y_{it} = 1|\alpha_t, \beta_t, \text{treat}_i) = \Pr(Y^*_{it} > 0|\beta_t, \text{treat}_i) = \Phi(\alpha_t + \beta_t \times \text{treat}_i), \quad t = 1, 2$$

$$\log(\tfrac{1+\rho_i}{1-\rho_i}) = \alpha_3 + \beta_3 \text{treat}_i.$$

نتایج این نوع مدل‌بندی پاسخ‌های دومتغیره با و بدون در نظر گرفتن ساختار همبستگی در جدول ۴.۶ نشان داده شده است، که تأییدی بر نتایج بر اساس جدول ۳.۶ است.

جدول ۴.٦: براورد پارامترهای مدل حاشیه‌ای با استفاده از مدل‌بندی با و بدون در نظر گرفتن ساختار همبستگی برای داده‌های بی‌خوابی (همبستگی ۲: همبستگی‌ای که به متغیرهای کمکی وابسته است، همبستگی ۱: همبستگی‌ای که به متغیرهای کمکی وابسته نیست)

پارامتر	با در نظر گرفتن همبستگی ۲		با در نظر گرفتن همبستگی ۱		بدون در نظر گرفتن همبستگی	
	براورد	انحراف معیار	براورد	انحراف معیار	براورد	انحراف معیار
$\alpha_۱$	۰/۵۷۳	۰/۱۰٦	۰/۵۸۸	۰/۱۰۲	۰/۵۷۳	۰/۱۲۱
$\alpha_۲$	۰/۰۰۰	۰/۰۰۰	−۰/۰۰۴	۰/۰۷۹	۰/۰۰۰	۰/۱۱۴
$\alpha_۳$	۱/۹۳۹	۰/۴۰۱	۱/۲۱۸	۰/۲۵۳	—	—
$\beta_۱$: درمان (داروی مؤثر)	۰/۰۴۳	۰/۱٦۲	۰/۰۰۷	۰/۱۲۳	۰/۰۴۳	۰/۱۷۳
$\beta_۲$: درمان (داروی مؤثر)	−۰/٦٦۸	۰/۱۲۵	−۰/٦۴۲	۰/۱۴۲	−۰/٦٦۸	۰/۱٦۹
$\beta_۳$: درمان (داروی مؤثر)	−۱/۴۱۲	۰/۵۳۴	—	—	—	—

۲.۳.۵.٦ مدل اثرهای تصادفی

یک دیدگاه اساسی که زیربنای مدل‌های اثرهای تصادفی است، این است که یک ناهمگنی طبیعی در میان افراد در ضرایب رگرسیونی آن‌ها وجود دارد، که این ناهمگنی می‌تواند از طریق توزیع احتمال بیان شود. همبستگی بین مشاهدات برای افراد یکسان، از متغیر مشترک غیر قابل مشاهده‌ی τ_i نشئت می‌گیرد. در این بخش به تشریح مدل لوجیت آمیخته‌ی خطی تعمیم‌یافته برای داده‌های دومتغیره‌ی بی‌خوابی می‌پردازیم.

فرض کنید مجموعه‌ی داده‌های ما شامل دو متغیر پاسخ دودویی برای هر فرد است، که این دو متغیر پاسخ برای iامین فرد، مقدار ۰ یا ۱ را اخذ می‌کنند. مدل اثرهای تصادفی به‌صورت زیر است:

$$\text{logit}[\Pr(Y_{it} = ۱|X_i, \tau_i)] = \alpha + \sum_{k=۱}^{K} \beta_k X_{itk} + \tau_i, \qquad t = ۱, ۲$$

که در آن τ_iها متغیرهای مستقل نرمال با میانگین ۰ و واریانس $\sigma^۲$ هستند. صورت مدل در نظر گرفته‌شده برای داده‌های بی‌خوابی به‌صورت زیر است:

$$\text{logit}[\Pr(Y_{i۱} = ۱|\text{treat}_i, \tau_i)] = \alpha + \beta_۱ \times \text{treat}_i + \tau_i,$$

$$\text{logit}[\Pr(Y_{i۲} = ۱|\text{treat}_i, \tau_i)] = \alpha + \beta_۱ \times \text{treat}_i + \alpha_۱ + \beta_۲ \times \text{treat}_i + \tau_i,$$

که در آن τ_i اثرهای تصادفی، α عرض از مبدأ در زمان اول، و $\alpha_۱$ اثر زمان بر لگاریتم بخت پاسخ در زمان دوم است. با این تعریف، $\alpha + \alpha_۱$ عرض از مبدأ برای مدل پاسخ در زمان دوم است. اثر درمان در زمان اول به‌وسیله‌ی $\beta_۱$، و اثر متقابل بین زمان و درمان به‌وسیله‌ی $\beta_۲$ نشان داده می‌شود. بنا بر این $\beta_۱ + \beta_۲$ اثر درمان در زمان دوم را نشان می‌دهد. روش ماکسیمم درست‌نمایی برای براورد کردن پارامترهای این مدل، مورد استفاده قرار گرفته است. تابع درست‌نمایی این مدل در اگرستی و ناتاراجان (۲۰۰۱) داده شده است. نتایج جدول ۵.٦ به‌وسیله‌ی نرم‌افزار STATA به دست آمده است.

جدول ۵.۶: براورد پارامترهای مدل اثرهای تصادفی برای داده‌های بی‌خوابی

	براورد به روش اثرهای تصادفی	
پارامتر	براورد	انحراف معیار
α	۱/۵۱۲	۰/۳۵۸
α_1: زمان	−۱/۵۲۱	۰/۳۹۲
β_1: درمان (داروی مؤثر)	۰/۰۵۱	۰/۴۵۱
β_2: درمان (داروی مؤثر)	−۱/۷۳۹	۰/۵۴۵
σ	۳/۷۵۲	۱/۴۳۷

نتایج این جدول نشان می‌دهد که اثر معناداری از نوع درمان روی احتمال پاسخ در زمان اول وجود ندارد، در حالی که اثر معنادار نوع درمان در پاسخ زمان دوم مشاهده می‌شود. همچنین براورد پارامتر σ در مدل اثرهای تصادفی، نشان‌دهنده‌ی همبستگی قوی بین پاسخ‌ها است.

۶.۵.۳.۳ مدل انتقالی

در مدل انتقالی مورد استفاده در این مسئله، اولین سؤال مورد نظر، در خصوص میزان تغییر در درمان است؛ به این معنا که انتقال بین زمان‌های متوالی چگونه صورت می‌گیرد؟ یک روش مناسب برای پاسخ دادن به این سؤال علمی، استفاده از مدل انتقالی است. در این‌جا از مدل انتقالی مرتبه‌ی اول برای پاسخ‌های دودویی تکراری استفاده کرده‌ایم. صورت کلّی مدل برای دو پاسخ به‌صورت زیر است:

$$\operatorname{logit}[\Pr(Y_{i1} = 1|\alpha, \beta)] = \alpha + \textstyle\sum_{k=1}^{K} \beta_k X_{i1k},$$

$$\operatorname{logit}[\Pr(Y_{i2} = 1|Y_{i1} = a; \alpha_a, \beta_a)] = \alpha_a + \textstyle\sum_{k=1}^{K} \beta_{ak} X_{i2k}, \qquad a = 0, 1$$

که در آن α و α_0 و α_1 پارامترهای عرض از مبدأ هستند و $\beta = (\beta_1, \dots, \beta_K)$ و به‌ازای $a = 0, 1$، $\beta_a = (\beta_{a1}, \dots, \beta_{aK})$ بردار پارامترها برای متغیرهای کمکی هستند. برای آشنایی با مباحث کلّی مدل‌های انتقالی و تابع درست‌نمایی برای یافتن براوردها به‌روش ماکسیمم درست‌نمایی، به گنجعلی و رضایی (۲۰۰۷) مراجعه شود. صورت مدل انتقالی برای داده‌های بی‌خوابی به‌صورت زیر است:

$$\operatorname{logit}[\Pr(Y_{i1} = 1|\alpha, \beta, \mathrm{treat}_i)] = \alpha + \beta \times \mathrm{treat}_i,$$

$$\operatorname{logit}[\Pr(Y_{i2} = 1|Y_{i1} = a; \alpha_a, \beta_a, \mathrm{treat}_i)] = \alpha_a + \beta_a \times \mathrm{treat}_i, \qquad a = 0, 1$$

که در آن α و α_a عرض از مبدأ هستند، β اثر درمان در زمان اول است و β_a اثر درمان در زمان دوم است. براورد پارامترهای مدل انتقالی در جدول ۶.۶ نشان داده شده است.

جدول ۶.۶: براورد پارامترهای مدل انتقالی برای داده‌های بی‌خوابی

پارامتر	Y_1		$Y_2 \mid Y_1 < ۳۰$		$Y_2 \mid Y_1 \geq ۳۰$	
	براورد	انحراف معیار	براورد	انحراف معیار	براورد	انحراف معیار
α	۰/۹۲۸	۰/۲۰۳	—	—	—	—
α_0	—	—	−۲/۰۱۵	۰/۵۳۲	—	—
α_1	—	—	—	—	۰/۶۲۴	۰/۲۲۶
β: درمان (داروی مؤثر)	۰/۰۷۲	۰/۲۸۹	—	—	—	—
β_0: درمان (داروی مؤثر)	—	—	۰/۳۲۹	۰/۷۲۱	—	—
β_1: درمان (داروی مؤثر)	—	—	—	—	−۱/۵۳۲	۰/۳۲۸

با استفاده از مدل انتقالی و این نتایج، اکنون بینش بهتری نسبت به فرایند تولید این داده‌ها داریم. زمانی که پاسخ اولیه کمتر از ۳۰ دقیقه است، اثر معناداری از دارو مشاهده نمی‌شود، اما زمانی که پاسخ زمان اول، بیش‌تر از ۳۰ دقیقه است، اثر مثبتی از دارو روی احتمال پاسخ در زمان دوم مشاهده می‌شود. این به این معنا است که دارو با احتمال کمتری برای بیمارانی که فاصله‌ی قبل از خواب آن‌ها کمتر از ۳۰ دقیقه است مؤثر خواهد بود و بنا بر این آگاهی داشتن از پاسخ زمان اول ممکن است پزشکان را از تجویز یک داروی غیر مؤثر منصرف کند.

۶.۶ تمرین‌ها

۱− به نظر شما پژوهشگر بر چه اساسی باید مدل طولی خود را معرفی کند؟

۲− برای مدل‌های حاشیه‌ای، اثرهای تصادفی و انتقالی معرفی‌شده در داده‌های بی‌خوابی (بخش ۳.۵.۶)، تابع درست‌نمایی را معرفی کنید.

۳− در مدل اثرهای تصادفی معرفی‌شده در مثال کاربردی داده‌های بی‌خوابی، اگر به‌جای پاسخ گسسته‌ی دودویی از مدل متغیر پنهان استفاده کنیم که در آن، متغیر پنهان پیوسته‌ی Y^*_{it} مشاهدات دودویی Y_{it} را تولید کند، و همچنین اگر واریانس اثر تصادفی τ_i که معرف واریانس بین آزمودنی‌ها است، σ^2 باشد و تغییرات داخل آزمودنی با $\frac{\pi^2}{3}$ نمایش داده شود

(واریانس توزیع لوژستیک یا خطای مدل Y_{it}^*)، نشان دهید همبستگی بین دو متغیر پنهان زمان اول و دوم (Y_{i1}^* و Y_{i2}^*) برابر است با $\rho = \frac{\sigma^2}{\sigma^2+\frac{\pi^2}{3}}$.

۴- اگر مدل اثرهای تصادفی معرفی‌شده در مثال داده‌های بی‌خوابی را به‌جای مدل اثر تصادفی‌ای که فقط عرض از مبدأ دارد، به مدل اثر تصادفی به‌صورت شیب و عرض از مبدأ تصادفی تعمیم دهیم به‌گونه‌ای که اثر تصادفی برای متغیر نوع درمان نیز در نظر گرفته شود،

آ) τ_i جدید را معرفی کنید و بر اساس آن به معرفی مدل جدید بپردازید. همچنین تعبیر هر دو نوع اثر تصادفی را ارائه کنید.

ب) کوواریانس بین Y_{i1}^* و Y_{i2}^* را محاسبه کنید.

پ) احتمال‌های حاشیه‌ای و توأم متغیرهای تصادفی دودویی را ارائه کنید.

ت) ضریب همبستگی پی‌یرسون و نسبت بخت‌های بین متغیرهای پاسخ را معرفی و محاسبه کنید.

ج) بر اساس نتایج ستون اول جدول ۳.۶ (با در نظر گرفتن همبستگی ۲)، نشان دهید که همبستگی بین پاسخ‌ها برای سطوح مختلف درمان، متفاوت است.

۵- در استفاده از مدل‌های انتقالی، چه راهی برای تعیین مرتبه‌ی مدل انتقالی پیشنهاد می‌کنید؟ در داده‌های آلودگی هوا بر سلامت کودکان، مدل انتقالی خود را معرفی و پیشنهاد خود در مورد مرتبه‌ی مدل را ارائه کنید.

۶- داده‌های آلودگی هوا بر سلامت کودکان را در نظر گرفته، با معرفی یک مدل حاشیه‌ای و استفاده از روش GEE پارامترهای مدل خود را براورد کنید. نتایج را تعبیر آماری کنید. از چه ساختاری برای ماتریس همبستگی استفاده کرده‌اید؟ توضیح دهید که چرا چنین انتخابی برای ساختار همبستگی داشته‌اید؟

۷- در مثال داده‌های مربوط به آسم، P_{i00} و P_{i01} و P_{i10} و P_{i11} چگونه به دست آمده‌اند؟