فصل ۳

مطالعات مقطعی: داده‌های دارای پاسخ ترتیبی

در این فصل به معرفی و بررسی مطالعات مقطعی با پاسخ تک‌متغیره‌ی ترتیبی می‌پردازیم. برای این منظور، در بخش اول، مفهومی از مطالعات مقطعی برای پاسخ‌های ترتیبی، همراه با نمادگذاری در این داده‌ها بیان می‌شود. در بخش دوم، توزیع مورد استفاده برای این نوع داده‌ها معرفی می‌شود. مدل‌های تجمعی برای داده‌های دارای پاسخ ترتیبی، با استفاده از مفهوم متغیر پنهان، و تعمیم مدل‌های تجمعی را در بخش سوم ارائه می‌کنیم. در بخش چهارم روش‌های محاسباتی برای براورد پارامترهای مدل‌های معرفی‌شده ارائه می‌شود. مثال‌های کاربردی با پاسخ ترتیبی مقطعی، در بخش پنجم آمده است. در بخش ششم، نیکویی برازش مدل‌های ترتیبی بیان می‌شود. در بخش هفتم، چگونگی استفاده از نرم‌افزار R برای برازش مدل با پاسخ ترتیبی آورده شده است.

۱.۳ مفاهیم و نمادگذاری

همان‌طور که در فصل‌های قبلی ذکر شد، مطالعات مقطعی تک‌متغیره، آن دسته از مطالعاتی هستند که در آن‌ها فقط یک متغیر پاسخ در مقطعی معلوم از زمان، مورد بررسی قرار می‌گیرد و می‌توان آن را به‌وسیله‌ی یک بردار $x = (x_1, \ldots, x_p)'$ از متغیرهای کمکی تشریح و تعبیر کرد. اغلب متغیرهای پاسخی که بیش از دو سطح دارند، ترتیبی‌اند و در نظر نگرفتن این ترتیب، از دست دادن قسمتی از اطلاعات نمونه را در پی خواهد داشت و این مسئله در نمونه‌های کوچک، مضرتر خواهد بود. در این فصل فرض خواهیم کرد که متغیر پاسخ، ترتیبی و مقادیر ممکن آن ۱، ۲، ...، J باشد.
به‌عنوان مثال، متغیر پاسخ می‌تواند رده‌ی فقر برای خانوار باشد، که شامل سطوح فقیر، شبه‌فقیر، غیر فقیر و شبه‌غنی یا غنی است. این متغیر را می‌توان به‌صورت «ترتیبی» در نظر گرفت و تأثیر متغیرهای کمکی، مانند وضع سواد سرپرست خانوار، وضع فعالیت سرپرست خانوار، بُعد خانوار و وضعیت تأهل سرپرست خانوار بر روی آن را بررسی کرد.
متغیر ترتیبی ممکن است از مکانیسم‌های مختلفی ایجاد شده باشد. این متغیر ممکن است از عملیات گروه‌بندی یک متغیر پیوسته ایجاد شده باشد. مثلاً در مثال رده‌ی فقر، پاسخ ترتیبیِ به دست آمده ممکن است حاصل از گروه‌بندی متغیر هزینه‌ی کل خانوار بر اساس شاخص فقر باشد. گروه دیگر متغیرهای ترتیبی، هنگامی به وجود می‌آید که پاسخگو با تحلیل مقدار نامعلومی از اطلاعات، قضاوتی را در مورد رتبه‌ی مقیاس ترتیبی در نظر بگیرد. مثلاً در مطالعه‌ی رضایت شغلی، هر فرد ممکن است بر اساس خصوصیات خود و محل کارش قضاوتی را در مورد این رضایت داشته باشد.

۲.۳ توزیع مناسب برای داده‌های مقطعی دارای پاسخ ترتیبی

فرض کنید متغیر تصادفی Y_i یکی از مقادیر گسسته‌ی ۱، ۲، ...، J را اخذ کند. قرار دهید

$$\pi_{ij} = \Pr(Y_i = j),$$

که نشان‌دهنده‌ی احتمال این است که i امین فرد، j امین رده را اخذ کند. فرض کنید که رده‌های پاسخ دوبه‌دو ناسازگار باشند؛ به این معنا که پاسخ هر فرد به متغیر تصادفی Y_i یکی از مقادیر ۱، ۲، ...، J را اخذ کند و هیچ فردی نمی‌تواند دو مقدار از این مقادیر را با هم اخذ کند. بنا بر این برای هر i داریم $\sum_{j=1}^{J} \pi_{ij} = 1$.

برای داده‌های گروه‌بندی‌شده، بررسی متغیرهای کمکی که معرف تعداد پاسخ‌ها در رده‌های مختلف پاسخ‌اند، مناسب خواهد بود. فرض کنید n_i تعداد افراد در گروه iام باشد و Y_{ij} تعداد پاسخ‌های گروه iام در رده‌ی jام با مقدار مشاهده‌شده‌ی y_{ij}. شایان ذکر است که $\sum_{j=1}^{J} y_{ij} = n_i$.

برای داده‌هایی که به‌صورت فردی ثبت شده‌اند، $n_i = 1$ است. در این حالت می‌توان پاسخ فرد iام را به‌صورت بردار $Y_i^* = (Y_{i1}^*, \dots, Y_{iJ}^*)$ نوشت که در آن Y_{ij}^* یک متغیر دودویی است که مقدار ۱ را زمانی اخذ می‌کند که iامین پاسخ در jامین رده قرار گیرد و در غیر این صورت مقدار صفر را اخذ می‌کند، و $\sum_{j=1}^{J} y_{ij}^* = 1$ چرا که یک و تنها یکی از پاسخ‌های دودویی Y_{ij}^* می‌تواند برای هر مورد، مقدار ۱ را اخذ کند.

در حالت گروه‌بندی، احتمال توزیع $Y_i = (Y_{i1}, \dots, Y_{iJ})$ به‌شرط معلوم بودن n_i با استفاده از توزیع چندجمله‌ای به‌صورت زیر مشخص می‌شود:

$$\Pr(Y_{i1} = y_{i1}, \dots, Y_{iJ} = y_{iJ}) = \binom{n_i}{y_{i1}, \dots, y_{iJ}} \pi_{i1}^{y_{i1}} \cdots \pi_{iJ}^{y_{iJ}}.$$

حالت خاص، زمانی است که $J = 2$ باشد و فقط دو رده‌ی پاسخ وجود داشته باشد، که در این حالت با توزیع دوجمله‌ای مواجه هستیم. برای اثبات، قرار دهید $y_{i1} = y_i$، $y_{i2} = n_i - y_i$، $\pi_{i1} = \pi_i$ و $\pi_{i2} = 1 - \pi_{i1}$.

۳.۳ مدل‌های مختلف برای تحلیل داده‌های مقطعی دارای پاسخ ترتیبی

در این بخش به معرفی و بررسی مدل‌های گوناگون برای مطالعات مقطعی با پاسخ‌های ترتیبی می‌پردازیم. مدل‌هایی با عبارت‌هایی که منعکس‌کننده‌ی مشخصه‌های ترتیبی باشند

(مانند روند یکنوا)، باعث بهبود امساک مدل می‌شوند. در این بخش به معرفی مدل‌های تجمعی معروف می‌پردازیم که برای پاسخ‌های ترتیبی مورد استفاده قرار می‌گیرند.

۱.۳.۳ مدل لوجیت‌های تجمعی

یک روش برای در نظر گرفتن ترتیب در اندازه‌گیری متغیر پاسخ، تشکیل احتمال‌های تجمعی به‌صورت

$$\Pr(Y \leq j|x) = \pi_1(x) + \cdots + \pi_j(x), \qquad j = 1, \ldots, J \tag{۱.۳}$$

است که با استفاده از آن‌ها، لوجیت‌های تجمعی به‌صورت زیر تعریف می‌شوند:

$$\begin{aligned} logit\Pr(Y \leq j|x) &= \log \frac{\Pr(Y \leq j|x)}{1 - \Pr(Y \leq j|x)} \\ &= \log \frac{\pi_1(x) + \cdots + \pi_j(x)}{\pi_{j+1}(x) + \cdots + \pi_J(x)}, \quad j = 1, \ldots, J-1 \end{aligned} \tag{۲.۳}$$

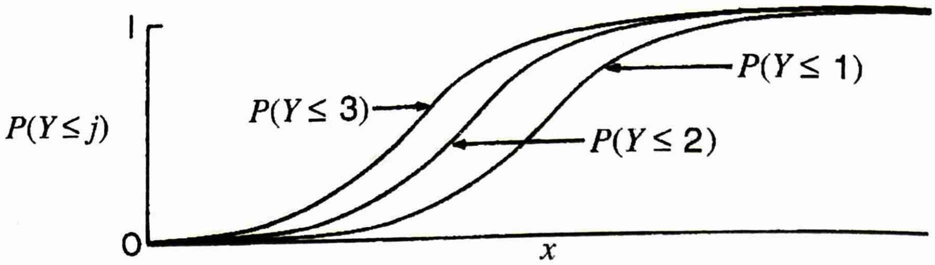
که در آن $\pi_j(x) = \Pr(Y = j|x)$. توجه داشته باشید که در هر سطح متغیر پاسخ، مدل لوجیت جداگانه‌ای در نظر گرفته می‌شود ($J - 1$ مدل لوجیت).

۲.۳.۳ مدل بخت متناسب

مدل بخت متناسب، مدلی است که به‌طور همزمان از همه‌ی لوجیت‌های تجمعی استفاده می‌کند به‌گونه‌ای که هر لوجیت تجمعی، مقدار ثابت خودش را دارد. این مدل عبارت است از

$$logit[\Pr(Y \leq j|x)] = \alpha_j + x'\beta, \qquad j = 1, \ldots, J-1 \tag{۳.۳}$$

که در آن، عبارت $\{\alpha_j\}$ یک تابع افزایشی از j است؛ زیرا $\Pr(Y \leq j|x)$ برای هر مقدار ثابت x با افزایش j افزایش می‌یابد و پیوند لوجیت هم یک تابع افزایشی از این احتمال است. α_jها پارامترهای نقاط برش نامیده می‌شوند. در این مدل، اثر x برای همه‌ی لوجیت‌ها یکسان (با ضریب β) در نظر گرفته می‌شود. شکل ۱.۳ بخت متناسب را برای $J = 4$ نمایش داده است. برای دو مقدار متفاوت بردار x، (مثلاً x_1 و x_2)، مدل لوجیت



شکل ۱.۳: مدل بخت متناسب برای پاسخی با ۴ رده

تجمعی (۳.۳) در رابطه‌ی زیر صدق می‌کند:

$$\operatorname{logit}[\Pr(Y \leq j|x_۱)] - \operatorname{logit}[\Pr(Y \leq j|x_۲)] = \log \frac{\frac{\Pr(Y\leq j|x_۱)}{\Pr(Y>j|x_۱)}}{\frac{\Pr(Y\leq j|x_۲)}{\Pr(Y>j|x_۲)}} = (x_۱ - x_۲)\beta,$$

که این لگاریتم، لگاریتم نسبت بخت‌های احتمال‌های تجمعی نامیده می‌شود. برای $x = x_۱$، بخت‌های پاسخ‌های کوچک‌تر یا مساوی مقدار j، به میزان $\exp[(x_۱ - x_۲)\beta]$ برابر همان بخت‌ها برای $x = x_۲$ است. لگاریتم نسبت بخت تجمعی، متناسب با اختلاف بین $x = x_۱$ و $x = x_۲$ است و از آن‌جا که این نسبت به j وابسته نیست، مکولا (۱۹۸۰) رابطه‌ی (۳.۳) را مدل بخت متناسب نام‌گذاری کرده است.

پس از معرفی مدل بخت متناسب، به بررسی آزمون استقلال شرطی با استفاده از این مدل می‌پردازیم. به‌منظور بررسی آزمون استقلال شرطی برای پاسخ ترتیبی Y، از مدل لوجیت تجمعی استفاده می‌کنیم، اما توابع ربط دیگر به آزمون‌های مشابهی منجر می‌شوند. فرض کنید Z یک متغیر کمکی اسمی باشد. در پی آنیم که $\Pr(Y \leq j|X, Z) = \Pr(Y \leq j|Z)$، یعنی استقلال X و Y به‌شرط Z را آزمون کنیم. اگر متغیر X ترتیبی یا پیوسته باشد، می‌توان از مدل زیر استفاده کرد:

$$\operatorname{logit}[\Pr(Y \leq j|X, Z = k)] = \alpha_j + \beta X + \beta_k^z, \quad k = ۱, \dots, K$$

که در آن β تأثیر متغیر ترتیبی یا پیوسته‌ی X را نشان می‌دهد و β_k^z تأثیر سطح kام متغیر اسمی Z با K رده (با محدودیت $\beta_K^z = ۰$) برای لوجیت‌های تجمعی است. برای استقلال شرطی کافی است $H_۰ : \beta = ۰$ را در مقابل $H_۱ : \beta \neq ۰$ آزمون کنیم. آزمون نسبت

درست‌نمایی، امتیاز یا آماره‌ی والد می‌تواند برای این منظور به کار رود.

اگر X متغیر اسمی باشد، مدل شرطی

$$\text{logit}[\Pr(Y \leq j|X = i, Z = k)] = \alpha_j + \beta_i^x + \beta_k^z, \quad i = ۱, \ldots, I, \quad k = ۱, \ldots, K$$

با محدودیت $\beta_I^x = ۰$ و $\beta_K^z = ۰$ می‌تواند برای آزمون استقلال شرطی به کار رود. اگر گواهی کافی بر رد فرض $H_۰ : \beta_۱^x = \cdots = \beta_I^x$ به دست نیاید، می‌توان فرض استقلال شرطی X و Y (به‌شرط Z) را پذیرفت.

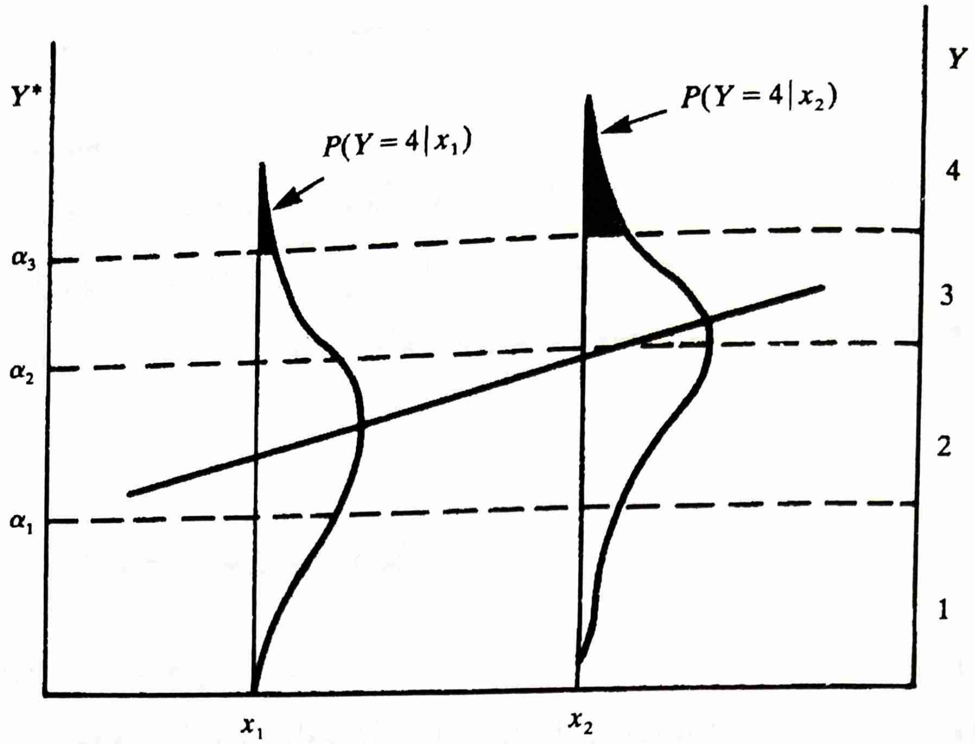
۳.۳.۳ مدل‌سازی پنهان

فرض کنید متغیر پیوسته‌ی مشاهده‌نشده‌ی Y^* اساس ایجاد متغیر گسسته‌ی Y باشد. به چنین متغیر مشاهده‌نشده‌ای متغیر پنهان گفته می‌شود. فرض کنید این متغیر دارای تابع توزیع تجمعی $\mathrm{G}(y^* - \eta)$ باشد که در آن y^* که حول پیشگوی خطی η تغییر می‌کند، وابسته به x است به‌گونه‌ای که $\eta(x) = x'\beta$. فرض کنید اگر $\alpha_{j-۱} < Y^* \leq \alpha_j$ باشد، $Y = j$ است؛ به این معنا که متغیر Y مقدار j را زمانی اخذ می‌کند که متغیر پنهان در jامین فاصله‌ی مقادیر خود قرار گیرد. بنا بر این

$$\Pr(Y \leq j|\ x) = \Pr(Y^* \leq \alpha_j|\ x) = \mathrm{G}(\alpha_j - x'\beta)$$

مدلی است که برای Y در نظر گرفته شده است و اشاره به تابع ربط $\mathrm{G}^{-۱}$ دارد که وارون تابع توزیع تجمعی Y^* است. در این مدل برای $j = ۱, \ldots, J - ۲$ داریم $\alpha_j \leq \alpha_{j+۱}$. به‌عنوان مثال در شکل ۲.۳، متغیر Y مقدار ۳ را زمانی اخذ می‌کند که $\alpha_۲ < Y^* \leq \alpha_۳$ باشد.

اگر $Y^* = x'\beta + \epsilon$ باشد به‌گونه‌ای که تابع توزیع تجمعی G برای ϵ توزیع لوژستیک باشد، $\mathrm{G}^{-۱}$ تابع ربط لوجیت است. همچنین در نظر گرفتن توزیع نرمال برای ϵ یک تابع ربط پروبیت برای احتمال‌های تجمعی ایجاد می‌کند. صرف نظر از این‌که نقاط برش $\{\alpha_j\}$، چگونه مقیاس را برای متغیر پنهان تغییر می‌دهند، برای پارامترهای مانا نسبت به سطح پاسخ، از β برای بررسی اثرهای روی Y استفاده می‌شود. این به آن معنا است که اثر پارامترها نسبت به انتخاب رده‌های Y پایا است. این ویژگی، امکان مقایسه‌ی براوردها از



شکل ۲.۳: اندازه‌های ترتیبی و مدل رگرسیونی برای یک متغیر پنهان

مطالعاتی را فراهم می‌کند که مقیاس‌های پاسخ متفاوتی دارند. اشاره به این نکته حائز اهمیت است که استفاده از تابع توزیع تجمعی به‌شکل $G(y^* - \eta)$ برای متغیر پنهان، منجر به پیش‌بینی $\alpha_j - x'\beta$ به‌جای $\alpha_j + x'\beta$ می‌شود. زمانی که $\beta > 0$ باشد، با افزایش x و در نظر گرفتن توزیع لوژستیک برای G، هر لوجیت تجمعی کاهش می‌یابد و در نتیجه احتمال‌های تجمعی نیز کاهش می‌یابند و به‌تبع آن جرم احتمال کمتر در مقادیر کمتر Y اتفاق می‌افتد. همچنین احتمال‌های تجمعی، در مقادیر بزرگ x، مقدار بزرگ‌تری دارند.

۴.۳.۳ مدل بخت رده‌ی مجاور

بخت‌های رده‌ی مجاور که با نماد $O(\text{adj}) = \frac{\Pr(Y=j)}{\Pr(Y=j+1)}$ نمایش داده می‌شود، معمولاً هنگامی مورد استفاده قرار می‌گیرد که احتمالات هر رده‌ی پاسخ، مورد علاقه باشد؛ یعنی زمانی که متغیر ترتیبی، رده‌بندی‌شده‌ی ذاتی باشد. مثالی از این نوع، وقتی است که متغیر میزان رضایت شغلی افراد، مورد بررسی است.

۵.۳.۳ مدل بخت نسبتی تسلسلی

بخت‌های تسلسلی به دو صورت $O(\mathrm{con1}) = \frac{\Pr(Y=j)}{\Pr(Y\geq j+1)}$ و $O(\mathrm{con2}) = \frac{\Pr(Y\leq j)}{\Pr(Y=j+1)}$ هستند و مربوط به فرایندهای تصمیم‌گیری دنباله‌ای‌اند که در آن‌ها، گزینه‌ها یا از کم به زیاد (نوع اول) یا از زیاد به کم (نوع دوم) سنجیده می‌شوند؛ مانند بررسی وضعیت تومور سرطانی که از کوچک تا خیلی بزرگ قابل رده‌بندی است و این اندازه در هر مرحله به‌طور دنباله‌ای تغییر وضعیت می‌دهد.

۶.۳.۳ مدل‌های پیچیده‌ی دیگر

مدل‌های لوجیت تجمعی دیگری نیز در رگرسیون لوژستیک ترتیبی به کار می‌روند. این مدل‌ها نیاز به مجموعه‌ای از پارامترهای ثابت به‌جای یک پارامتر ثابت دارند. مدل‌های این بخش، فرض یکسان بودن اثرهای بخت‌های متناسب را برای لوجیت‌های تجمعی مختلف در نظر گرفته‌اند. یک مزیت این روش، سادگی تفسیر و تحلیل اثرها است، چرا که فقط به یک بردار پارامتری برای هر پیشگو نیاز است. در مدل‌های پیچیده، این مدل‌های ساده با در نظر گرفتن اثرهای مجزا از طریق جایگزین کردن β با β_j در مدل (۳.۳) تعمیم می‌یابد.

۷.۳.۳ مدل‌های تجمعی با تابع توزیع تجمعی به‌عنوان تابع ربط

مدل‌های لوجیت تجمعی از تابع ربط لوجیت استفاده می‌کنند. همان‌طور که در مدل‌های خطی تعمیم‌یافته‌ی یک‌متغیره، تابع‌های ربط دیگری قابل استفاده بودند، در این حالت نیز می‌توان از تابع‌های ربط دیگری استفاده کرد. فرض کنید G^{-1} معرف تابع ربطی است که وارون تابع توزیع تجمعی پیوسته‌ی G است. مدل ربط تجمعی

$$G^{-1}[\Pr(Y \leq j|\, x)] = \alpha_j + x'\beta, \tag{۴.۳}$$

احتمال‌های تجمعی را به پیشگوی خطی مرتبط می‌کند. در حالت خاص، تابع ربط لوجیت $G^{-1}(u) = \log[\frac{u}{1-u}]$ وارون تابع توزیع تجمعی لوژستیک استاندارد است. مشابه مدل بخت متناسب (۳.۳)، اثرهای x در مدل (۴.۳) برای همه‌ی نقاط برش به‌ازای

$j = ۱, ۲, \ldots, J - ۱$ مشابه و یکسان فرض شده‌اند.

استفاده از تابع توزیع نرمال استاندارد Φ برای G منجر به مدل پروبیت تجمعی می‌شود. این تابع ربط، هنگامی مناسب است که توزیع متغیر پنهان Y^* نرمال باشد. پارامترها در مدل پروبیت بر حسب متغیر پنهان Y^* تفسیر می‌شوند. به‌عنوان مثال، در مدل $\Phi^{-۱}[\Pr(Y \leq j)] = \alpha_j - x'\beta$ چون $Y^* = x'\beta + \epsilon$ برای $\epsilon \sim \mathrm{N}(۰, ۱)$ دارای توزیع تجمعی نرمال (Φ) است، β (در صورت وجودِ تنها یک متغیر کمکی) این‌گونه تفسیر می‌شود که با یک واحد افزایش در x، $E(Y^*)$ نیز به اندازه‌ی β واحد افزایش می‌یابد. مدل‌های لوجیت تجمعی، برازشی مشابه با مدل‌های پروبیت تجمعی دارند، ولی تفسیر پارامترها در این مدل‌ها با استفاده از نسبت بخت احتمال‌های تجمعی، ساده‌تر است.

توزیع مقادیر فرین برای Y^* منجر به مدلی به‌صورت

$$\log\{-\log[۱ - \Pr(Y \leq j|\, x)]\} = \alpha_j + x'\beta$$

می‌شود که به تابع ربط لگ–لگ مکمل برای داده‌های ترتیبی‌ای معروف است. مدل ترتیبی که از این تابع ربط استفاده کند، اغلب به مدل خطرهای متناسب معروف است؛ زیرا این مدل از تعمیم مدل خطر متناسب برای داده‌های بقا نتیجه گرفته شده است (پرنتیس و گلوکلر، ۱۹۷۸). این مدل دارای خاصیت زیر است:

$$\Pr(Y > j|\, x_۱) = [\Pr(Y > j|\, x_۲)]^{\exp((x_۱ - x_۲)\beta)}.$$

با این تابع ربط، به‌عنوان مثال با یک پیشگوی خطی ساده، با افزایش x، مقدار $\Pr(Y \leq j)$ به عدد ۱ سریع‌تر از صفر همگرا است. تابع ربط لگ–لگ به‌صورت $\log\{\log[\Pr(Y \leq j)]\}$ زمانی پیشنهاد می‌شود که $\Pr(Y \leq j)$ به عدد صفر سریع‌تر از یک همگرا است.

۴.۳ روش‌های براورد

۱.۴.۳ روش شبه‌درست‌نمایی

روش شبه‌درست‌نمایی که برای پاسخ‌های ترتیبی به کار می‌رود، مانند روش شبه‌درست‌نمایی فصل قبل برای پاسخ‌های دودویی مقطعی است، با این تفاوت که پاسخ‌های ترتیبی و توزیع پاسخ‌ها در این بخش، چندجمله‌ای است. به عبارت دیگر، معادلات براوردگر برای پاسخ‌های ترتیبی، همان معادلات براوردگر بخش قبل است و فقط توزیع متغیرهای پاسخ، متفاوت است، که این اختلاف، منجر به تغییر میانگین و واریانس در معادلات براوردگر تعمیم‌یافته می‌شود. بنا بر این از تکرار آن در این بخش خودداری می‌کنیم. در نرم‌افزارهای موجود مانند SAS، با تبدیل پاسخ‌های ترتیبی با J رده به $J-۱$ پاسخ دودویی (همان‌طور که در بخش ۲.۳ توضیح داده شد)، و با در نظر گرفتن همبستگی بین این پاسخ‌های دودویی در معادلات براوردگر، پارامترهای مدل براورد می‌شوند.

۲.۴.۳ معادلات درست‌نمایی و براوردگرهای ماکسیمم درست‌نمایی

فرض کنید که β برداری از پارامترهای مدل باشد و α برداری از پارامترهای برش (آستانه). با توجه به مدل‌های ارائه‌شده در بخش‌های قبل (مدل ۳.۳)، داریم:

$$\Pr(Y_i = m|x_i, \beta, \alpha) = F(\alpha_m + x_i'\beta) - F(\alpha_{m-۱} + x_i'\beta).$$

بنا بر این، احتمال مشاهده‌ی $Y = j$ برای iامین فرد $(j = ۱, ۲, \ldots, J)$ برابر است با $\Pr(Y_i = j|x_i, \beta, \alpha)$.

اگر مشاهدات مستقل باشند، معادله‌ی درست‌نمایی به‌صورت زیر بیان می‌شود:

$$\begin{aligned} L(\beta, \alpha|y, X) &= \prod_{i=۱}^{n} \Pr(Y_i = y_i|x_i, \beta, \alpha) \\ &= \prod_{i=۱}^{n} [F(\alpha_{y_i} + x_i'\beta) - F(\alpha_{y_i-۱} + x_i'\beta)]. \end{aligned} \tag{۵.۳}$$

با گرفتن لگاریتم از دو طرف معادله‌ی فوق داریم:

$$\ell(\beta, \alpha|y, X) = \sum_{i=۱}^{n} \ln[F(\alpha_{y_i} + x_i'\beta) - F(\alpha_{y_i-۱} + x_i'\beta)]. \tag{۶.۳}$$

این معادله می‌تواند از طریق نرم‌افزارهای موجود مانند SAS و SPSS بهینه شود و براورد پارامترهای مدل از آن به دست آید. این معادله همچنین می‌تواند از طریق روش‌های عددی برای براورد پارامترها ماکسیمم شود. مادالا (۱۹۸۳، صص. ۴۴۹—۴۴۸) گرادیان و ماتریس هسی را برای استفاده از روش نیوتون—رافسون ارائه کرده است. همچنین نشان داده شده است که روش نیوتون—رافسون به ماکسیمم فراموضعی همگرا است. براوردهای به دست آمده سازگار، به‌طور تقریبی نرمال، و به‌طور تقریبی کارا هستند.

۵.۳ مثال‌های کاربردی

در این بخش به معرفی سه مثال کاربردی با پاسخ ترتیبی می‌پردازیم. مثال اول به بررسی وضعیت فقر افراد خانوار با استفاده از داده‌های طرح آمارگیری از هزینه و درامد خانوارها می‌پردازد، مثال دوم در باره‌ی طول مدت بیکاری افراد خانوار بر اساس داده‌های طرح آمارگیری نیروی کار مرکز آمار ایران است، و مثال آخر، بررسی نوع وسایل ضد بارداری مورد استفاده‌ی زنان کشور السالوادور بر اساس آمارگیری نفوس و سلامت آن کشور می‌باشد.

۱.۵.۳ مقایسه‌ی فقر با استفاده از داده‌های طرح آمارگیری از هزینه و درامد خانوارها — سال ۱۳۸۴

در این بخش، از داده‌های طرح آمارگیری از هزینه و درامد خانوارها استفاده خواهیم کرد که در بخش اول به معرفی آن پرداخته شد. طرح آمارگیری از هزینه و درامد خانوارهای شهری و روستایی، یکی از طرح‌های آماری مهم و گسترده در کشور است که با قدمتی ۴۶ ساله توسط مرکز آمار ایران اجرا می‌شود. به‌علت وسعت پوشش مکانی و نیز نوع اقلام مورد پرسش، نتایج حاصل از این طرح، بسیاری از اطلاعات مورد نیاز اقتصادی،

همچون شناسایی الگوی مصرف، توزیع درامد، و نیز میزان برخورداری خانوارها از امکانات و تسهیلات زندگی اجتماعی را در مقیاس داخلی و بین‌المللی فراهم می آورد، که این اطلاعات در سیاست‌گذاری‌های کلان اقتصادی دولت و بخش خصوصی، مفید واقع می‌شوند.

حال از یک مدل لوژستیک تجمعی ترتیبی برای مدل‌بندی داده‌های تشریح‌شده در زیربخش ۱.۲.۲.۱ فصل ۱ به‌صورت زیر استفاده می‌کنیم (می‌توان متغیر پنهان ایجادکننده‌ی متغیر ترتیبی را همان هزینه‌ی کل دانست که بر اساس شاخص‌های فقر، رده‌بندی شده است):

$$\log\left[\frac{\Pr(Y_i \leq j|x_i)}{\Pr(Y_i > j|x_i)}\right] = \alpha_j - x_i'\beta, \qquad j = ۱, ۲, ۳, ۴ \tag{۷.۳}$$

تابع درست‌نمایی مدل به‌صورت زیر قابل بیان است:

$$L(\alpha, \beta|y, x) = \prod_{i=۱}^{n}\left[\frac{\exp(\alpha_{y_i} - x_i'\beta)}{۱ + \exp(\alpha_{y_i} - x_i'\beta)} - \frac{\exp(\alpha_{y_i-۱} - x_i'\beta)}{۱ + \exp(\alpha_{y_i-۱} - x_i'\beta)}\right].$$

مقادیر براورد پارامترها، بر اساس تابع درست‌نمایی معرفی‌شده، در جدول ۱.۳ ارائه شده است. نتایج این جدول نشان می‌دهد که در صورتِ ثابت فرض کردن همه‌ی متغیرها به‌غیر از جنس، بخت سطوح پایین فقر (نسبت به سطوح بالا) برای زنان سرپرست خانوار، بیش‌تر از مردان سرپرست خانوار است. به‌طریق مشابه، بخت سطوح پایین فقر برای بی‌سوادان، بیش‌تر از باسوادان است. بخت سطوح پایین فقر برای شاغلان، کمتر از بیکاران جویای کار است. بخت سطوح پایین فقر برای سرپرستان دارای همسر، بیش‌تر از هرگز ازدواج نکرده‌ها می‌باشد. بخت سطوح پایین فقر برای دارندگان مسکن، کمتر از دیگران است. و بالاخره بخت سطوح پایین فقر، تابعی صعودی از بعد خانوار است؛ به این معنا که هرچه بعد خانوار بیش‌تر می‌شود، احتمال فقیر بودن یا شبه‌فقیر بودن افراد خانوار، افزایش می‌یابد.

جدول ۱.۳: براورد پارامترهای مدل لوژستیک تجمعی برای پاسخ ترتیبی وضعیت فقر خانوار (طرح آمارگیری از هزینه و درامد خانوارها– ۱۳۸۴)

	پارامتر	براورد	انحراف معیار	P–مقدار
نقطه‌ی برش	α_1	۰/۵۶۲–	۰/۱۷۱	۰/۰۰۱
	α_2	۱/۲۱۶	۰/۱۷۲	۰/۰۰۰
	α_3	۲/۴۹۱	۰/۱۷۳	۰/۰۰۰
جنس	مرد (مبنا)	–	–	–
	زن	۰/۴۵۸–	۰/۱۱۷	۰/۰۰۰
وضع سواد	باسواد(مبنا)	–	–	–
	بی‌سواد	۱/۳۷۶–	۰/۰۵۰	۰/۰۰۰
وضع فعالیت	سایر(مبنا)	–	–	–
	شاغل	۰/۱۱۶	۰/۱۰۸	۰/۲۸۴
	بیکار جویای کار	۰/۸۳۰–	۰/۱۶۹	۰/۰۰۰
	دانشجوی بیکار	۰/۱۴۵	۰/۱۰۹	۰/۱۸۴
وضع زناشویی	هرگز ازدواج نکرده (مبنا)	–	–	–
	دارای همسر	۰/۱۶۳–	۰/۱۳۰	۰/۲۰۸
	بی‌همسر بر اثر طلاق	۰/۲۶۶	۰/۱۶۲	۰/۱۰۰
	بی‌همسر بر اثر فوت	۰/۱۷۷–	۰/۲۴۲	۰/۴۶۴
وضع مسکن	سایر (مبنا)	–	–	–
	مالک	۰/۵۴۳	۰/۰۶۲	۰/۰۰۰
	مستأجر	۰/۰۷۱	۰/۰۶۸	۰/۲۹۲
بعد خانوار	بیش‌تر از شش نفر (مبنا)	–	–	–
	یک نفر	۲/۸۵۷	۰/۱۱۷	۰/۰۰۰
	دو نفر	۲/۲۳۶	۰/۰۶۶	۰/۰۰۰
	سه نفر	۱/۶۴۶	۰/۰۵۶	۰/۰۰۰
	چهار نفر	۱/۲۱۴	۰/۰۵۲	۰/۰۰۰
	پنج نفر	۰/۷۵۰	۰/۰۵۵	۰/۰۰۰

۲.۵.۳ بیکاری در طرح آمارگیری نیروی کار سال ۱۳۸۵

در این بخش، مدل تک‌متغیره‌ی ترتیبی برای داده‌های طرح آمارگیری نیروی کار بهار ۱۳۸۵ معرفی می‌شود. متغیر پاسخ، طول مدت بیکاری افراد در بهار ۱۳۸۵ است که در زیربخش ۲.۲.۲.۱ به‌تفصیل بیان شده است. متغیرهای کمکی پیوسته و گسسته‌ای که در این مدل به کار می‌روند و کدهای مربوط به آن‌ها در جدول ۲.۳ بیان شده‌اند. مدل‌های لوجیت تک‌متغیره‌ای که آن‌ها را در این بخش معرفی کردیم، برای داده‌های مورد استفاده در این قسمت، با در نظر گرفتن متغیرهای کمکی معرفی‌شده در جدول ۲.۳، برازش داده می‌شوند.

داده‌هایی که از این آمارگیری برای هر عضو خانوارهای نمونه گردآوری شده است، به‌روش مناسبی وزن‌دهی می‌شوند. این وزن‌ها پس از اِعمال سه مرحله‌ی وزن‌دهی «محاسبه‌ی وزن پایه»، «تعدیل وزن برای واحدهای بی‌پاسخ» و «تعدیل وزن بر اساس پیش‌بینی‌های جمعیتی» (دویل و سارندال، ۱۹۹۲) به دست آمده‌اند، که از آن‌ها در

بخش‌های بعدی و در تابع درست‌نمایی برای براورد پارامترها استفاده می‌شود. در مدل‌های مورد استفاده برای پاسخ ترتیبی، آن دسته از متغیرهای کمکی که بیش از دو سطح داشته باشند (مانند سطح تحصیلات و وضع زناشویی)، به‌عنوان متغیرهای عاملی در نظر گرفته و با تعریف معمول به متغیرهای دودویی تبدیل می‌شوند.

جدول ۲.۳: متغیرهای کمکی و کدهای مربوط به آن‌ها

متغیر کمکی	کد مربوط
سن	متغیر پیوسته
بستگی با سرپرست خانوار	۱: سرپرست خانوار ۲: بقیه‌ی اعضای خانوار
وضع مهاجرت	۱: مهاجر ۲: غیر مهاجر
وضع تحصیلات	۱: در حال تحصیل بودن ۲: در حال تحصیل نبودن
محل زندگی	۱: شهری ۲: روستایی
جنس	۱: مرد ۲: زن
سطح تحصیلات	۱: زیر دیپلم ۲: دیپلم یا پیش‌دانشگاهی ۳: تحصیلات دانشگاهی
وضع زناشویی	۱: ازدواج کرده ۲: طلاق‌گرفته یا بیوه‌شده ۳: هرگز ازدواج نکرده
تجربه‌ی کار قبلی	۱: تجربه داشتن ۲: تجربه نداشتن
وضع فعالیت قبلی	۱: شاغل ۲: در حال تحصیل ۳: بقیه‌ی موارد
روش جستجوی کار	۱: جستجو نکردن ۲: پرس و جو از دوستان و آشنایان ۳: جستجوی منابع مالی برای شروع فعالیت خوداشتغالی یا تقاضای جواز کسب ۴: درج آگهی در روزنامه یا مطالعه‌ی آگهی‌های استخدام ۵: بقیه‌ی موارد
تعداد افراد شاغل در خانوار	۱: صفر ۲: یک ۳: بیش از دو نفر
تعداد افراد در خانوار	۱: یک یا دو نفر ۲: سه یا چهار نفر ۳: بیش از چهار نفر

اکنون با استفاده از مدل تک‌متغیره‌ی ترتیبی به تحلیل داده‌های نیروی کار بهار سال ۱۳۸۵ می‌پردازیم. مدل مورد استفاده برای پاسخ طول مدت بیکاری، مدل لوجیت تجمعی حاشیه‌ای است که در آن اثرهای متغیرهای کمکی روی لگاریتم بخت انتخاب

رده‌های کوچک‌تر نسبت به رده‌های بزرگ‌تر به‌صورت زیر براورد می‌شود:

$$\log[\frac{\Pr(Y_i \leq j;\alpha,\beta)}{\Pr(Y_i > j;\alpha,\beta)}] = \alpha_j + x_i'\beta, \qquad j = ۱,\ldots,J-۱$$

که در آن Y_i پاسخ kامین فرد، x_i بردار متغیرهای کمکی با مؤلفه‌ی x_{ik} متناظر با k امین متغیر کمکی kامین فرد، و J تعداد رده‌های ترتیبی متغیرهای پاسخ است. α_jها نیز به‌ازای $j = ۱,\ldots,J-۱$، نقاط برش هستند و نشان‌دهنده‌ی لگاریتم بخت انتخاب رده‌های کوچک‌تر نسبت به رده‌های بزرگ‌ترند، هنگامی که متغیرهای کمکی، برابر با صفر در نظر گرفته شده باشند. $\alpha = (\alpha_۱,\ldots,\alpha_{J-۱})'$ بردار پارامترهای آستانه‌ای است، که $\alpha_۱ \leq \alpha_۲ \leq \cdots \leq \alpha_{J-۱}$ است و β بردار ضرایب رگرسیونی برای متغیرهای کمکی است.
احتمال این‌که پاسخ فرد kام در اولین سطح پاسخ $(j = ۱)$ باشد، به‌صورت زیر است:

$$\pi_{i۱} = \Pr(Y_i = ۱|\alpha,\beta) = \frac{\exp(\alpha_۱ + x_i'\beta)}{۱ + \exp(\alpha_۱ + x_i'\beta)},$$

و برای $j = ۲,\ldots,J-۱$ داریم:

$$\begin{aligned}\pi_{ij} &= \Pr(Y_i = j|\alpha,\beta)\\ &= \Pr(Y_i \leq j|\alpha,\beta) - \Pr(Y_i \leq j-۱|\alpha,\beta)\\ &= \frac{\exp(\alpha_j + x_i'\beta)}{\exp(\alpha_j + x_i'\beta)} - \frac{\exp(\alpha_{j-۱} + x_i'\beta)}{۱ + \exp(\alpha_{j-۱} + x_i'\beta)}.\end{aligned}$$

تابع درست‌نمایی و لگاریتم تابع درست‌نمایی برای این مدل با توجه به در نظر گرفتن وزن هر فرد، به‌ترتیب به‌صورت زیرند:

$$L = \prod_{i=۱}^{n} \pi_{i,y_i}{}^{W_i}$$

و

$$\ell(\theta) = \sum_{i=۱}^{n} W_i \log(\pi_{i,y_i}),$$

که در این روابط، n تعداد افراد در مطالعه، θ برداری حاوی همه‌ی پارامترهای مدل، و W_i وزن نمونه‌گیری برای kامین فرد است که در آن π_{i,y_i} احتمال پاسخ ترتیبی فرد kام است، و $\pi_{i,J} = ۱ - \sum_{l_۱=۱}^{J-۱} \pi_{i,l_۱}$. با توجه به این‌که در تابع درست‌نمایی، وزن هر یک از افراد

نمونه در نظر گرفته شده است، نتایج مدل‌بندی، قابل تعمیم به کل جامعه می‌باشد. شاید بهتر باشد که تابع درست‌نمایی معرفی‌شده را تابع شبه‌درست‌نمایی بنامیم، زیرا وزن‌های «فردی» در آن در نظر گرفته شده است.

جدول ۳.۳: براورد پارامترها و انحراف معیار در مدل لوجیت تجمعی برای طول مدت بیکاری بر اساس داده‌های نیروی کار (مقادیری که باعلامت * مشخص شده‌اند، در سطح ۵ درصد معنادارند)

پارامتر	کل جمعیت		مردان		زنان	
	براورد	انحراف معیار	براورد	انحراف معیار	براورد	انحراف معیار
$\alpha_{۱,۱}$	−۰/۱۹۴۵	۰/۳۳۳	۰/۲۹۲	۰/۴۲۹	۰/۶۱۰	۰/۶۷۳
$\alpha_{۱,۲}$	۰/۸۲۰*	۰/۳۲۹	۱/۳۸۳*	۰/۴۳۱	۱/۳۵۷*	۰/۶۹۰
$\alpha_{۱,۳}$	۱/۸۶۹*	۰/۳۳۷	۲/۳۱۸*	۰/۴۳۷	۲/۸۱۴*	۰/۶۹۹
$\alpha_{۱,۴}$	۲/۹۶۳*	۰/۳۳۵	۳/۳۷۷*	۰/۴۳۳	۴/۰۱۵*	۰/۷۱۳
سن	−۰/۰۶۳*	۰/۰۰۷	−۰/۰۵۶*	۰/۰۰۸	−۰/۱۱۴*	۰/۰۲۰
بستگی با سرپرست خانوار						
مبنا: بقیه‌ی اعضای خانوار	—	—	—	—	—	—
سرپرست خانوار	۰/۷۹۱*	۰/۱۹۹	۰/۵۰۰*	۰/۲۳۱	۰/۵۷۲	۰/۶۶۵
وضع مهاجرت						
مبنا: غیر مهاجر	—	—	—	—	—	—
مهاجر	۰/۶۰۶*	۰/۱۹۱	۰/۵۴۸*	۰/۲۲۶	۰/۹۴۸*	۰/۳۵۲
وضع تحصیلات						
مبنا: در حال تحصیل نبودن	—	—	—	—	—	—
در حال تحصیل بودن	۰/۷۷۳*	۰/۲۷۱	۰/۸۵۳*	۰/۳۵۹	۰/۶۴۳	۰/۴۱۱
جنس						
مبنا: زن	—	—	—	—	—	—
مرد	۰/۲۹۷*	۰/۱۲۳	—	—	—	—
سطح تحصیلات						
مبنا: تحصیلات دانشگاهی	—	—	—	—	—	—
زیر دیپلم	−۰/۰۵۷	۰/۱۴۷	−۰/۱۷۴	۰/۱۹۸	−۰/۱۶۶	۰/۳۰۱
دیپلم یا پیش‌دانشگاهی	−۰/۳۴۸*	۰/۱۳۶	−۰/۵۱۰*	۰/۱۹۷	−۰/۳۹۵	۰/۲۰۲
وضع زناشویی						
مبنا: هرگز ازدواج نکرده	—	—	—	—	—	—
ازدواج کرده	۰/۲۲۴	۰/۱۴۰	۰/۴۴۷*	۰/۲۰۰	۰/۰۴۹	۰/۲۰۸
طلاق‌گرفته یا بیوه‌شده	۰/۳۶۱	۰/۳۸۲	۰/۹۶۶	۰/۵۷۹	۰/۳۵۸	۰/۵۹۵
تجربه‌ی کار قبلی						
مبنا: تجربه نداشتن	—	—	—	—	—	—
تجربه داشتن	۰/۵۴۱*	۰/۱۸۴	۰/۶۸۸*	۰/۲۳۳	۰/۴۵۶	۰/۳۰۴
وضع فعالیت قبلی						
مبنا: بقیه‌ی موارد	—	—	—	—	—	—
شاغل	۰/۴۰۴*	۰/۱۹۵	۰/۴۱۵	۰/۲۳۸	۰/۲۶۹	۰/۳۷۵
در حال تحصیل	−۰/۴۰۲*	۰/۱۷۹	−۰/۳۲۰	۰/۲۴۴	−۰/۸۵۹*	۰/۳۰۷
روش جستجوی کار						
مبنا: بقیه‌ی موارد	—	—	—	—	—	—
جستجو نکردن	۰/۴۶۰	۰/۲۸۵	۰/۰۱۴	۰/۳۳۴	۱/۳۳۱*	۰/۶۲۶
پرس و جو از دوستان و آشنایان	−۰/۳۰۱	۰/۲۰۰	−۰/۶۹۴*	۰/۲۵۶	۰/۴۰۰	۰/۳۴۹
شروع خوداشتغالی یا تقاضای جواز کسب	−۰/۷۷۰*	۰/۲۳۹	−۱/۱۱۶*	۰/۲۹۸	−۰/۱۰۵	۰/۴۳۳
درج یا مطالعه‌ی آگهی در روزنامه	−۰/۸۳۴*	۰/۲۰۴	−۱/۳۳۸*	۰/۲۷۰	−۰/۰۷۴	۰/۳۳۸
$-۲ \log L_{Null}$	۳۳۸۵۱۲۶/۰		۲۴۰۹۸۹۵/۴		۸۹۷۳۸۶/۸	
$-۲ \log L_{Full}$	۳۱۴۷۱۴۹/۶		۲۲۷۷۳۵۹/۴		۸۳۵۰۳۲/۱	

بنا بر این جدول ۳.۳ نتایج مدل‌بندی حاشیه‌ای طول مدت بیکاری برای کل جامعه و به‌طور مجزا به تفکیک جنس را بیان می‌کند. نتایج این جدول نشان می‌دهد که هرچه سن بیش‌تر باشد، طول مدت بیکاری با احتمال بیش‌تری افزایش می‌یابد. همچنین نتایج نشان می‌دهد احتمال این‌که سرپرست خانوار، مدت بیکاری کوتاه‌تری نسبت به بقیه‌ی اعضای خانوار داشته باشد، بیش‌تر است. چنین تعبیری برای افرادی که مهاجرند و کسانی که در حال تحصیل هستند نیز برقرار است. همچنین مردها با احتمال بیش‌تری مدت بیکاری

کوتاه‌تر دارند. اثر سطح تحصیلات نشان می‌دهد که افراد در سطح تحصیلات دیپلم یا پیش‌دانشگاهی، مدت بیکاری طولانی‌تری دارند. نتایج همچنین نشان می‌دهد مردانی که ازدواج کرده‌اند، مدت بیکاری کوتاه‌تری دارند. جدول ۳.۳ همچنین نشان می‌دهد افراد یا مردانی که تجربه‌ی کار قبلی دارند، مدت بیکاری کوتاه‌تری دارند و احتمال کوتاه بودن مدت بیکاری برای افرادی که قبل از بیکاری‌شان شاغل بوده‌اند نسبت به بقیه‌ی افراد (خانه‌دار، بازنشسته و ...) بیش‌تر است. این احتمال برای افرادی که قبل از بیکاری‌شان تحصیل می‌کرده‌اند، کمتر است. این اثر برای زنان و کل افراد نیز معنادار است. با توجه به تعریف تعدیل‌شده‌ی بیکاری (سازمان بین‌المللی کار[۱]، ۱۹۹۲) نتایج نشان می‌دهد زنانی که به‌دلایل موجه در جستجوی کار نبوده‌اند (به‌دلیل این‌که یا کاری یافته‌اند و قرار است آن را دیرتر شروع کنند، یا این‌که در انتظار بازگشت به شغل قبلی خود هستند) مدت بیکاری کوتاه‌تری نسبت به زنانی دارند که در مؤسسات کاریابی ثبت‌نام کرده‌اند یا با کارفرما تماس داشته‌اند یا روش‌های دیگری را برای جستجوی کار انتخاب کرده‌اند. مردانی که به دنبال منابع مالی و امکاناتی برای شروع فعالیت خوداشتغالی هستند یا تقاضای جواز کسب یا پروانه‌ی کار دارند، مدت بیکاری طولانی‌تری نسبت به بقیه‌ی افراد دارند، که این به‌دلیل طولانی‌تر بودن فرایند آغاز این نوع کارها است. این نتایج برای مردان و کل جمعیت مورد مطالعه که از طریق درج آگهی در روزنامه‌ها یا مطالعه‌ی آگهی‌های استخدام به دنبال کار می‌گردند، نسبت به بقیه‌ی افراد نیز برقرار است. در جدول ۳.۳، $-2\log L_{\text{Null}}$ منفی دو برابر لگاریتم درست‌نمایی بدون در نظر گرفتن هیچ متغیر کمکی است و $-2\log L_{\text{Full}}$ منفی دو برابر لگاریتم درست‌نمایی با در نظر گرفتن همه‌ی متغیرهای کمکی در مدل است.

۳.۵.۳ استفاده از وسایل ضد بارداری برای زنان

داده‌هایی که در این زیربخش، مورد بررسی قرار می‌دهیم، در آمارگیری نفوس و سلامت کشور السالوادور در سال ۱۹۸۵ گردآوری شده‌اند.[۲] این داده‌ها مربوط به نوع وسیله‌ی ضد بارداری مورد استفاده‌ی ۳۱۶۵ زن است که در السالوادور ازدواج کرده‌اند. زنان ازدواج‌کرده بر اساس فاصله‌های سنی ۵ ساله رده‌بندی شده‌اند. همچنین نوع وسیله‌ی ضد

---

[۱] International Labour Organization (ILO)

[۲] *http://www.stat.ufl.edu/~jgill/contraception.dat*

بارداری مورد استفاده‌ی زنان به سه دسته‌ی «سترون‌سازی»، «روش‌های دیگر» و «هیچ روش» طبقه‌بندی شده است. از آن‌جا که در سترون‌سازی شانسی برای بچه‌دار شدن وجود ندارد، و در استفاده از روش‌های دیگر هنوز شانس کمی برای بچه‌دار شدن وجود دارد، و بدون استفاده از روش‌های ضد بارداری، شانس زیادی وجود دارد، نوع وسیله‌ی ضد بارداری می‌تواند به‌عنوان یک متغیر ترتیبی در نظر گرفته شود. داده‌های این مثال در جدول ۴.۳ ارائه شده است و مدل‌بندی و تحلیل نتایج این داده‌ها در بخش ۷.۳ آمده است.

جدول ۴.۳: داده‌های مربوط به استفاده از وسیله‌ی ضد بارداری برای زنانی که در کشور السالوادور ازدواج کرده‌اند

سن	سترون‌سازی	روش‌های دیگر	هیچ روش	تعداد کل
۱۵	۳	۶۱	۲۳۲	۲۹۶
۲۰	۸۰	۱۳۷	۴۰۰	۶۱۷
۲۵	۲۱۶	۱۳۱	۳۰۱	۶۴۸
۳۰	۲۶۸	۷۶	۲۰۳	۵۴۷
۳۵	۱۹۷	۵۰	۱۸۸	۴۳۵
۴۰	۱۵۰	۲۴	۱۶۴	۳۳۸
۴۵	۹۱	۱۰	۱۸۳	۲۸۴

۶.۳ نیکویی برازش مدل ترتیبی

قبل از وارد شدن به مبحث مربوط به محاسبه‌ی نتایج مربوط به مطالعات مقطعی با پاسخ ترتیبی با استفاده از نرم‌افزار R، به معرفی معیار نیکویی برازش مدل برای این نوع پاسخ‌ها می‌پردازیم؛ چرا که در نظر گرفتن نیکویی برازش مدل نیز یک مسئله‌ی مهم است که باید به آن پرداخته شود. برای پاسخ‌های ترتیبی، آزمون نیکویی برازش نسبتی (آزمون نسبت درست‌نمایی) ممکن است برای ارزیابی برازش مدل‌های معرفی‌شده برای پاسخ‌های ترتیبی مناسب نباشد. در بعضی شبیه‌سازی‌ها نشان داده شده است که آماره‌ی کیبش در هنگامی که داده‌ها خیلی پراکنده باشند، دارای توزیع خی‌دو نیست. به‌عنوان مثال، زمانی که تعداد مشاهدات در خانه‌هایی که از طریق رده‌بندی متغیرهای کمکی به دست آمده‌اند، کم باشد، آماره‌ی کیبش مناسب نیست (کوهلر، ۱۹۸۶؛ کوهلر و لارنتز، ۱۹۸۰؛ لارنتز، ۱۹۷۸). در داده‌هایی که ما در این فصل کتاب به کار برده‌ایم، به‌دلیل این‌که متغیرهای کمکی زیادی در نظر گرفته شده‌اند و اغلب آن‌ها گسسته‌اند، استفاده از آزمون نسبت درست‌نمایی پیشنهاد نمی‌شود. بنا بر این باید به دنبال معیار دیگری باشیم.

برای پاسخ‌های ترتیبی، دو نسخه‌ی تغییریافته‌ی ضریب همبستگی چندگانه‌ی R^2 وجود دارد. یک نسخه‌ی آن توسط کاکس و اسنل (۱۹۸۹) بسط داده شده است و دیگری توسط نگل‌کرک (۱۹۹۱) به‌گونه‌ای تغییر یافته است که بتواند مقادیر صفر تا یک را اخذ کند. شبه R^2 ناگل‌کرک که برای نیکویی برازش مدل لوجیت تجمعی به کار برده می‌شود، به‌صورت زیر تعریف می‌شود:

$$R^2 = \frac{1 - \exp[-2(n)^{-1}(\log L_{\text{Full}} - \log L_{\text{Null}})]}{1 - \exp[2(n)^{-1}\log L_{\text{Null}}]}$$

که در آن L_{Full} و L_{Null} به‌ترتیب، تابع درست‌نمایی مدل برازانده‌شده و مدل پوچ (مدلی که فقط مقادیر ثابت را در نظر می‌گیرد) هستند. شبه R^2 می‌تواند به‌عنوان نسبتی از واریانس تفسیر شود که توسط مدل تشریح می‌شود.

۷.۳ دستورهای R برای برازش مدل در داده‌های مقطعی با پاسخ ترتیبی

در این بخش با استفاده از دستورهای مورد استفاده در نرم‌افزار R به تحلیل داده‌های مربوط به زنان کشور السالوادور (مثال زیربخش ۳.۵.۳ و پیوست ۲.۲ را ببینید) می‌پردازیم. همان‌طور که در فصل قبل بیان شد، یکی از راه‌های فراخوانی داده‌ها در نرم‌افزار R، استفاده از دستور `read.table` است که می‌تواند داده‌هایی را فراخوانی کند که به‌صورت فایل txt ذخیره شده‌اند. اکنون روش دیگر فراخوانی داده‌ها از اینترنت را بیان می‌کنیم. در صورت وصل بودن به اینترنت، داده‌های جدول ۴.۳ با استفاده از دستورهای زیر از وب‌گاه (سایت) مربوط، قابل فراخوانی است. در صورتی که به اینترنت وصل نباشید می‌توانید با وارد کردن داده‌های جدول ۴.۳ از آن‌ها استفاده کنید. با توجه به این‌که به‌منظور استفاده از مدل‌های نسبت بخت‌ها برای پاسخ‌های ترتیبی، متغیر پاسخ و متغیرهای کمکی اسمی یا ترتیبی موجود در مدل باید به‌صورت عاملی باشند، از دستور `factor` برای این متغیرها استفاده شده است. چون متغیر سن به‌صورت رده‌بندی‌شده است، برای این‌که بتوان برای هر یک از رده‌های متغیر کمکی سن، پارامتری براورد کرد، باید از دستور `contrasts` استفاده کرد. اگر از `contrasts = FALSE` استفاده شود، به

ماتریسی همانی دست می‌یابیم که از روش معمول تبدیل یک متغیر عاملی با K سطح به $K-1$ متغیر دودویی با مبنای آخرین سطح متغیر استفاده می‌کند. انتخاب هر یک از این دو روش بر عهده‌ی تحلیلگر است. در زیر، دستورهای مربوط به فراخوانی داده‌ها از اینترنت بیان شده است.

```
contraception.mat = as.matrix(read.table("http://www.stat.ufl.edu
/~jgill/contraception.dat",header=TRUE))
contraception.df  = data.frame(expand.grid(Response=1:3,"Age"=
contraception.mat[,1]),"Freq"=as.numeric(t(contraception.mat[,2:4])))
contraception.df$Response = factor(contraception.df$Response)
levels(contraception.df$Response) = c("Sterilization","Other","None")
contraception.df$Age = factor(contraception.df$Age)
 levels(contraception.df$Age) = c("15-19","20-24","25-29",
"30-34","35-39","40-44","45-49")
contrasts(Age, contrasts = FALSE)
attach(contraception.df)
```

پس از فراخوانی داده‌ها و آماده‌سازی آن، نیاز به مدل‌بندی و انتخاب بهترین مدل است. مدل نسبت بخت‌ها با استفاده از تابع polr در کتابخانه‌ی MASS در نرم‌افزار R به داده‌های دارای پاسخ ترتیبی برازانده می‌شود، که برای داده‌های مربوط به استفاده از وسیله‌ی ضد بارداری با استفاده از دستورهای زیر به نتایج شکل ۳.۳ دست خواهیم یافت:

```
library(MASS)
library(nnet)
contraception.plo <- polr(Response ~ Age, weights=Freq)
summary(contraception.plo)
```

```
Re-fitting to get Hessian

Call:
polr(formula = Response ~ Age, weights = Freq)

Coefficients:
                Value      Std. Error    t value
Age1(β1)     1.2442203    0.12352385   10.072713
Age2(β2)     0.5522919    0.07886593    7.002922
Age3(β3)    -0.2615745    0.07294721   -3.585805
Age4(β4)    -0.7805545    0.07942790   -9.827207
Age5(β5)    -0.5719095    0.08712568   -6.564190
Age6(β6)    -0.4330931    0.09826667   -4.407325

Intercepts:
                            Value    Std. Error    t value
Sterilization|Other(α1)   -0.8810     0.0430      -20.4999
Other|None(α2)            -0.1728     0.0402       -4.2951

Residual Deviance: 5963.335
AIC: 5979.335
```

شکل ۳.۳: نتایج برازش مدل تجمعی ترتیبی برای داده‌های زنان کشور السالوادور

نتایج شکل ۳.۳ معنادار بودن متغیرهای کمکی و مقادیر ثابت را نشان می‌دهد. با توجه به این‌که در نرم‌افزار R، مدل معرفی‌شده به‌صورت $\operatorname{logit}(\Pr(Y_i \leq j|x)) = \alpha_j - x'\beta$ است، ضرایب منفی نشان‌دهنده‌ی بیش‌تر بودن احتمال مربوط به رده‌های پایین‌تر می‌باشد. بنا بر این، به‌عنوان مثال می‌توان گفت که گروه سنی ۳۴—۳۰ (Age4 در شکل ۳.۳) با احتمال بیش‌تری در گروه استفاده از وسایل ضد بارداری سترون‌سازی است و با کمترین احتمال به گروه استفاده از هیچ روش ضد بارداری تعلق دارند. معیار کیبش مانده‌ها که برابر با منفی دو برابر لگاریتم تابع درست‌نمایی است، مقدار ۵۹۶۳/۳۳۵ را نشان می‌دهد، که با توجه به بزرگ‌تر بودن آن نسبت به ۳۱۵۷ = ۸ − ۳۱۱۶۵ =d.f. انتظار داریم که مدل بتواند بهبود یابد. معیار آکائیکه که در شکل ۳.۳ با AIC نمایش داده شده است، برابر است با $۵۹۷۹/۳۳۵ = \text{مانده‌های کیبش} + ۲ \times p = -۲ \times \log L + ۲ \times p = \text{AIC}$ (که در آن $p = ۸$ تعداد پارامترهای مدلِ براورد شده است). همچنین با استفاده از نتایج شکل ۳.۳ می‌توان به‌عنوان مثال به محاسبه‌ی $\Pr(Y \leq ۲|\text{Age} = ۱)$ یا $\Pr(Y \leq ۲|\text{Age} = ۶)$ پرداخت. با توجه به این‌که برای تبدیل متغیر سن به ۶ متغیر جدید از دستور `contrasts(Age, contrasts = FALSE)` استفاده کرده‌ایم، احتمال‌های فوق برای گروه‌های سنی ۱۵—۱۹ و ۴۵—۴۹ به‌صورت زیر محاسبه می‌شوند:

$$\Pr(Y \leq ۲|\text{Age} = 1) = \operatorname{logit}(\alpha_2 - \beta_1),$$

$$[۱ + \exp(-(-۰/۱۷۲۸ - (۱)(۱/۲۴۴)))]^{-۱} = ۰/۱۹۵۲,$$

$$\Pr(Y \leq ۲|\text{Age} = ٦) = \text{logit}(\alpha_۲ - \beta_٦) = ۰/۵٦۵.$$

به‌منظور براورد کردن مدل پروبیت ترتیبی، می‌توان از دستور زیر استفاده کرد:

```
contraception.pro = polr(Response ~ Age, weights=Freq, data=
contraception.df, method ="probit")
```

شایان ذکر است که گزینه‌ی method در تابع polr امکان براورد پارامترهای مدل را بر اساس تابع‌های ربط مختلف، از جمله کوشی (cauchit)، لگ—لگ مکمل (cloglog)، پروبیت (probit) و لوجیت (logistic) فراهم می‌آورد که در آن، پیش‌فرض تابع ربط، لوجیت است.
همچنین می‌توان با استفاده از دستور predict یا fitted.values به برازش احتمال‌های رده‌های مختلف پاسخ بر اساس برازش مدل نسبت بخت‌ها دست یافت.

```
p.fit = predict(contraception.plo,type="probs")
p.fit = fitted.values(contraception.plo,type="probs")
```

با استفاده از مقادیر برازانده‌شده می‌توان نمودارهای مختلفی که مورد نظر محقق و تحلیلگر است، با به‌کارگیری دستورهای شکل در فصل قبل رسم کرد.

۸.۳ تمرین‌ها

۱— در جدول ۵.۳ که مربوط به تعداد افراد تصادف‌کرده در رده‌های مختلف پاسخ به تفکیک جنس، محل زندگی و استفاده یا عدم استفاده از کمربند ایمنی است، رده‌های پاسخ عبارت‌اند از: (۱) آسیب‌ندیده، (۲) آسیب‌ندیده و نیازی به خدمات

اورژانس پزشکی نیست، (۳) آسیب‌دیده و به خدمات اورژانس پزشکی منتقل شده، (۴)آسیب‌دیده، بستری‌شده در بیمارستان و زنده مانده، (۵) آسیب‌دیده و مرده. خروجی در شکل ۴.۳ که بر اساس مدل لوجیت تجمعی است، با استفاده از متغیرهای کمکی که همه‌ی آن‌ها دودویی‌اند، ارائه شده است.

آ) توضیح دهید که چرا ۴ نقطه‌ی آستانه‌ای وجود دارد؟ توضیح دهید که چگونه این چهار نقطه‌ی آستانه‌ای، توزیع پاسخ براورد شده برای مردان در نقاط شهری را که کمربند ایمنی می‌بندند، تعیین می‌کنند.

ب) یک فاصله‌ی اطمینان برای اثر جنسیت به‌شرط استفاده از کمربند ایمنی و منطقه‌ی زندگی بسازید و آن را تفسیر کنید.

پ) نسبت بخت‌های تجمعی براورد شده را بین متغیر پاسخ و استفاده از کمربند ایمنی برای افرادی که در منطقه‌ی روستایی زندگی می‌کنند و برای افرادی که در مناطق شهری زندگی می‌کنند به‌شرط جنسیت پیدا کنید. بر اساس آن توضیح دهید که چگونه اثر استفاده از کمربند ایمنی در مناطق تغییر می‌کند. همچنین براورد اثر متقابل (۰/۱۲۴۴−) را تفسیر کنید.

جدول ۵.۳: تعداد افراد تصادف‌کرده در رده‌های مختلف پاسخ به تفکیک جنس، محل زندگی و استفاده یا عدم استفاده از کمربند ایمنی

جنس	محل زندگی	استفاده از کمربند ایمنی	پاسخ				
			۱	۲	۳	۴	۵
زن	شهر	نه	۷۲۸۷	۱۷۵	۷۲۰	۹۱	۱۰
		بلی	۱۱۵۸۷	۱۲۶	۵۷۷	۴۸	۸
	روستا	نه	۳۲۴۶	۷۳	۷۱۰	۱۵۹	۳۱
		بلی	۶۱۳۴	۹۴	۵۶۴	۸۲	۱۷
مرد	شهر	نه	۱۰۳۸۱	۱۳۶	۵۶۶	۹۶	۱۴
		بلی	۱۰۹۶۹	۸۳	۲۵۹	۳۷	۱
	روستا	نه	۶۱۲۳	۱۴۱	۷۱۰	۱۸۸	۴۵
		بلی	۶۶۹۳	۷۴	۳۵۳	۷۴	۱۲

۲− با استفاده از مدل لوجیت تجمعی برای داده‌های جدول ۶.۳ که مربوط به وضعیت رضایت شغلی و درامد به تفکیک جنسیت است،

آ) اثر درامد براورد شده‌ی $\hat{\beta}_1 = -۰/۵۱$ را با براورد بعد از ادغام رده‌های پاسخ به سه رده از طریق ادغام رده‌های (i) خیلی راضی و به‌طور متوسط راضی، و (ii) خیلی ناراضی

Parameter			DF	Estimate	Std Error
Intercept1			1	3.3074	0.0351
Intercept2			1	3.4818	0.0355
Intercept3			1	5.3494	0.0470
Intercept4			1	7.2563	0.0914
gender	female		1	−0.5463	0.0272
gender	male		0	0.0000	0.0000
location	rural		1	−0.6988	0.0424
location	urban		0	0.0000	0.0000
seatbelt	no		1	−0.7602	0.0393
seatbelt	yes		0	0.0000	0.0000
location*seatbelt	rural	no	1	−0.1244	0.0548
location*seatbelt	rural	yes	0	0.0000	0.0000
location*seatbelt	urban	no	0	0.0000	0.0000
location*seatbelt	urban	yes	0	0.0000	0.0000

شکل ۴.۳: نتایج مدل‌بندی تمرین ۱

و کمی راضی مقایسه کنید. این مسئله چه خاصیتی از مدل را منعکس می‌کند؟

ب) $\frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$ را با استفاده از رده‌های اصلی با $\frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$ بر اساس ادغام رده‌ها بر اساس قسمت (آ) محاسبه و مقایسه کنید. آیا موافقید که یک ضعف ادغام رده‌های پاسخ ترتیبی این است که معناداری اثرها تقلیل می‌یابد؟

پ) آزمون کنید که آیا با در نظر گرفتن اثر متقابل بین درامد و جنس، نتایج مدل بهبود می‌یابد یا نه. این نکته را تفسیر کنید.

جدول ۶.۳: وضعیت رضایت شغلی و درامد به تفکیک جنس

جنس	درامد (دلار)	رضایت شغلی			
		خیلی ناراضی	کمی راضی	به‌طور متوسط راضی	خیلی راضی
زن	< ۵۰۰۰	۱	۳	۱۱	۲
	۱۵۰۰۰–۵۰۰۰	۲	۳	۱۷	۳
	۲۵۰۰۰–۱۵۰۰۰	۰	۷	۸	۵
	۲۵۰۰۰+	۰	۲	۴	۲
مرد	< ۵۰۰۰	۱	۱	۲	۱
	۱۵۰۰۰–۵۰۰۰	۰	۳	۵	۱
	۲۵۰۰۰–۱۵۰۰۰	۰	۰	۷	۳
	۲۵۰۰۰+	۰	۱	۹	۶

۳– جدول ۷.۳ اشاره به یک آزمایش بالینی برای درمان سرطان ریه دارد. در این مطالعه، بیماران به‌طور تصادفی به دو گروه درمانی طبقه‌بندی شده‌اند. در گروه درمانی دائمی (متوالی)، از ترکیب مشابهی از عامل‌های درمان دارویی در هر دوره‌ی درمان استفاده شده است. در درمان دیگر، سه ترکیب متفاوت به بیماران داده می‌شود که از یک

دوره به دوره‌ی دیگر متناوب است.

آ) یک مدل لوجیت تجمعی برای درمان و جنس برازش دهید و نتایج را تفسیر کنید.

ب) مدلی را که شامل یک عبارت اثر متقابل باشد، برازش کنید و آن را تفسیر کنید. آیا این برازش بهتر است؟ توضیح دهید که نتایج این مدل، هم‌ارز با نتایج مدلی است که از چهار ترکیب جنس و درمان به‌عنوان سطوح یک عامل واحد استفاده می‌کند.

۷.۳: پاسخ به درمان دارویی در هر رده از سطوح مختلف گروه درمانی و جنس

درمان	جنس	پاسخ به درمان دارویی			
		پیشرفت بیماری	بدون تغییر	بهبود اندک	بهبود کامل
متوالی	مرد	۲۸	۴۵	۲۹	۲۶
	زن	۳	۱۲	۵	۲
دوره به دوره	مرد	۴۱	۴۴	۲۰	۲۰
	زن	۱۲	۷	۳	۱

۴– برای یک جدول پیشایندی (I)–بعدی با یک متغیر پاسخ ترتیبی،(Y)، و یک متغیر اسمی،(X)، با I رده، نشان دهید که استقلال آماری بین دو متغیر، معادل با رابطه‌ی زیر است:

$$\text{logit}[\Pr(Y \leq j|X=i)] = \alpha_j, \qquad i = ۱,\ldots,I, \quad j = ۱,\ldots,J-۱.$$

۵– مدل زیر را برای یک پاسخ ترتیبی با ۳ رده و یک متغیر کمکی دودویی در نظر بگیرید.

$$\log\left[\frac{\Pr(Y_i \leq j|X)}{\Pr(Y_i > j|X)}\right] = \alpha_j + \beta_j X_i, \qquad i = ۱,\ldots,n, \quad j = ۱,۲.$$

بدون ابزار یارانه‌ای، چگونه می‌توان براورد پارامترهای $\alpha_۱$ و $\alpha_۲$ و $\beta_۱$ و $\beta_۲$ را به دست آورد؟ همچنین چگونه می‌توان فرض $H_۰ : \beta_۱ = \beta_۲$ را در مقابل $H_۱ : \beta_۱ \neq \beta_۲$ آزمون کرد؟