به‌کارگیری توابع ربط مناسب در مدل‌های خطی تعمیم‌یافته برای آنالیز داده‌های طولی با متغیر پاسخ رتبه‌ای

مریم واعظ هوس،[۱] دکتر غلامرضا بابایی،[۲] دکتر سقراط فقیه‌زاده[۳]

چکیده

هرگاه از هر عضو نمونه در چند موقعیت، مشاهداتی داشته باشیم، مشاهدات به دست آمده را اندازه‌گیری‌های مکرر می‌نامیم. داده‌های مکرری را که در آن‌ها موقعیت تکرار مشاهدات، نقاط زمانی هستند، داده‌های طولی و مطالعات از این نوع را مطالعات طولی می‌نامند.

مطالعات طولی که در آن‌ها متغیر پاسخ، از نوع رتبه‌ای (نظیر شدت درد، درجه‌ی بهبودی، شدت استفراغ و ...) باشد جایگاه خاصی در علوم پزشکی دارد، زیرا پزشک برای بررسی سیر و تأثیر روش‌های درمانی مختلف، بیمار را به طور مکرر در طول زمان مورد معاینه قرار می‌دهد.

با توجه به ماهیت داده‌ها در مطالعات طولی با پاسخ رتبه‌ای، در تحلیل این نوع داده‌ها باید تمهیدات خاصی در نظر گرفته شود، زیرا متغیر مورد اندازه‌گیری برای هر آزمودنی بیش از یک

---

[۱] *کارشناس ارشد آمار زیستی مرکز آمار ایران*

[۲] *دانشیار آمار زیستی دانشگاه تربیت مدرس*

[۳] *دانشیار آمار زیستی دانشگاه تربیت مدرس*

بار اندازه‌گیری می‌شود و لذا به علت عدم استقلال پاسخ‌ها در روش اندازه‌گیری مکرر نمی‌توان از روش آنالیز رگرسیون معمول استفاده نمود.

در این مقاله ضمن منظور نمودن همبستگی بین مشاهدات، ضرایب β از روش حداقل مربعات وزنی، با به‌کارگیری توابع ربط مناسب براورد شده است.

در پایان، روش مذکور به کمک نرم‌افزار SAS در یک مطالعه‌ی واقعی نشان داده شده است.

واژه‌های کلیدی: مطالعه‌ی طولی، متغیر پاسخ رتبه‌ای، مدل خطی تعمیم‌یافته، حداقل مربعات وزنی.

مقدمه

در بسیاری از تحقیقات پزشکی برای بررسی سیر و تأثیر روش‌های درمانی مختلف، پزشک بیمار را به طور مکرر در طول زمان مورد معاینه قرار می‌دهد. مشاهدات حاصل، اندازه‌گیری مکرر نامیده می‌شوند.

موقعیت‌هایی که اندازه‌گیری صورت می‌گیرد لزوماً زمان نیستند، مثل داده‌های مربوط به دو چشم هر فرد. داده‌های مکرری که موقعیت‌های تکرار مشاهدات، نقاط زمانی هستند داده‌های طولی و مطالعات از این نوع را مطالعات طولی می‌نامند.

این نوع مطالعات، تغییرات هر فرد در طول زمان را از اندازه‌ی سطح مبنای‌شان مشخص می‌کند که مطالعات مقطعی قادر به چنین کاری نیستند. از آن جا که مجموعه‌ی مشاهدات روی هر فرد با هم همبستگی دارند و این همبستگی باید در استنباط‌های معتبر، به حساب آورده شود لذا داده‌های حاصل از مطالعات طولی نیاز به روش‌های آماری ویژه دارند که در آن‌ها باید این همبستگی‌ها لحاظ شوند.

در پژوهش حاضر، حالتی مد نظر است که متغیر پاسخ به صورت متغیر رتبه‌ای در مدل ظاهر می‌شود. این مدل‌ها در تحقیقات مربوط به علم پزشکی قابل بهره‌برداری هستند، زیرا در این تحقیقات غالباً متغیر پاسخ، سیر بیماری و یا سیر درمان است و هر دو ویژگی بیماری و درمان که از متغیرهای کیفی هستند، ماهیت رتبه‌ای دارند.

مطالعات مشابه زیادی در این راستا انجام شده است که در ذیل به اختصار به برخی از آن‌ها اشاره می‌شود:

اگرستی [۱] در مقاله‌ای تحت عنوان «بررسی مدل‌هایی برای داده‌های پاسخ رتبه‌ای تکراری»، ضمن معرفی و بررسی مدل‌های مطلوب، با یک مثال عملی روی داده‌های واقعی، به برازش مدل‌های معرفی شده و براورد پارامترهای آن پرداخته است. در این مقاله ضمن معرفی توابع ربط مناسب برای پاسخ‌های رتبه‌ای و روش‌های براورد پارامترها، از توابع ربط لجیت تجمعی، لجیت رسته‌های مجاور و همچنین روش حداقل مربعات وزنی در یک مثال عملی استفاده شده است.

اگرستی و ناتارجان[۲] روش‌های مختلف مدل‌بندی متغیرهای پاسخ رتبه‌ای را وقتی داده‌ها از نوع خوشه‌ای هستند، ارائه می‌دهند که تعمیم کار پندرگست، جانگ، نیوتن و فیشر[۳] می‌باشد که برای داده‌های دو حالتی عرضه کردند. این مقاله به ارائه‌ی روش‌های آنالیز پاسخ‌های رتبه‌ای وقتی داده‌ها از نوع خوشه‌ای هستند از قبیل خوشه‌ی پاسخ‌های تکراری برای یک آزمودنی در یک مطالعه‌ی طولی می‌پردازد که در آن هر خوشه، یک مجموعه از اندازه‌گیری‌های تکراری روی یک آزمودنی است. در این راستا، دو نوع بررسی می‌شود: یکی مدل‌های حاشیه‌ای که توزیع حاشیه‌ای پاسخ را به صورت یک تابع از متغیرهای توضیحی، به طور جداگانه از همبستگی داخل آزمودنی، روی زمان مدل‌بندی می‌کند و دیگری مدل‌های آثار تصادفی که در آن، یک اثر تصادفی برای هر آزمونی معرفی می‌شود. در این مقاله، روش‌های معرفی شده، با یک مثال عملی نیز نشان داده شده است.

لیانگ و زیگر[۴] روش‌های مختلف تحلیل داده‌های طولی را مرور کرده، یک دسته‌بندی کلی از مدل‌های مورد استفاده در مطالعات طولی شامل مدل‌های حاشیه‌ای، مدل‌های تغییر وضعیتی (مارکف) و مدل‌های آثار تصادفی را ارائه کرده‌اند.

استارم، وی و ویر[۵] روش ماکسیمم درست‌نمایی[۱] و حداقل مربعات وزنی[۲] (WLS) را برای مدل‌بندی لجیت‌های تجمعی داده‌های اندازه‌گیری مکرر پیشنهاد کردند.

اشبی و دیگران[۶] خلاصه‌ی بیش از یکصد مقاله را که تا سال ۱۹۹۲ در خصوص

---

[۱] *Maximum Likelihood*

[۲] *Weighted Least-Squares*

روش‌های تحلیل داده‌های رسته‌ای همبسته به چاپ رسیده است جمع‌آوری کرده و برحسب ویژگی‌هایی نظیر نوع پاسخ و نوع متغیرها دسته‌بندی کرده‌اند که می‌تواند منبع مفیدی را برای مرور روش‌های مختلف تحلیل در مطالعات طولی ارائه نماید.

کنوارد و دیگران[۷] روش معادلات براورد تعمیم‌یافته را تحلیل داده‌های طولی با پاسخ‌های رتبه‌ای به کار گرفته و آن را با روش ماکسیمم درست‌نمایی مقایسه نموده‌اند.

علوی مجد[۸] روش حداقل مربعات وزنی را در تحلیل داده‌های طولی با پاسخ رتبه‌ای در حالت وجود داده‌های گم شده به کار گرفته است.

قربانی[۹] روش بیزی را در تحلیل داده‌های طولی رتبه‌ای در دو حالت وجود و عدم مقادیر گم شده به کار گرفته و در خاتمه با یک مثال عملی نتایج حاصل از این روش را با روش حداقل مربعات وزنی مورد مقایسه قرار داده است.

روش انجام کار

در مطالعات طولی که با اندازه‌گیری‌های مکرر از هر فرد تحت مطالعه مواجهیم، به دلیل این که هر فرد در چند نوبت مورد مشاهده قرار می‌گیرد و مشاهدات مربوط به هر فرد قاعدتاً دارای نوعی همبستگی می‌باشند، لذا در مدل‌بندی این پاسخ‌های همبسته لازم است ملاحظات خاصی را در نظر گرفت.

مدل‌بندی آماری پاسخ‌های همبسته در مطالعات طولی از دو دیدگاه زیر قابل دسترسی است:

۱ـ تعیین مدل رگرسیونی مناسب برای بیان ارتباط بین متغیرهای پاسخ و متغیرهای مستقل.

۲ـ توجه به همبستگی درونی بین پاسخ‌های مکرر.

در مدل‌بندی داده‌های طولی، از مدل‌های حاشیه‌ای، مدل‌های تصادفی و مدل‌های انتقالی (مارکوف) استفاده می‌شود. در این مقاله از مدل حاشیه‌ای در مدل‌بندی استفاده شده است و لذا به طور مختصر این مدل را معرفی می‌کنیم.

در یک مدل حاشیه‌ای، رگرسیون پاسخ بر روی متغیرهای توضیحی به صورت مجزا

از همبستگی درون عنصری مدل‌بندی می‌شود. ما امید حاشیه‌ای $E(y_{ij})$ را بـه صـورت تابعی از متغیرهای توضیحی مدل‌بندی می‌کنیم. امید ریاضی حاشیه‌ای عبارت اسـت از میانگین پاسخ برای زیر گروهی از جامعه که x آن‌ها یک مقدار خاص است.

یک مدل حاشیه‌ای با فرضیات ذیل تعریف می‌شود:

ـ امیـد ریاضـی حاشـیه‌ای متغیـر پاسـخ $E(y_{ij}) = \mu_{ij}$ بـه وسـیله‌ی $h(\mu_{ij}) = X'_{ij}\beta$ به متغیرهای توضیحی y_{ij} وابسته است که h یک تابع ربط مثل لجیت برای پاسخ‌های دوتایی و یا لگاریتم برای پاسخ‌های شمارشی است.

ـ واریانس حاشیه‌ای به صورت $\mathrm{var}(y_{ij}) = \mathrm{V}(\mu_{ij})\phi$ به میانگین حاشیه‌ای وابسته است که V یک تابع واریانس معلوم و ϕ یک پارامتر مقیاس می‌باشد که ممکن است نیاز به براورد آن باشد.

ـ همبستگی بین y_{ij} و y_{ik} تابعی از میانگین حاشیه‌ای و احتمالاً پارامترهـای اضـافی α است. یعنی

$$\mathrm{Corr}(\mathrm{y}_{\mathrm{ij}}, \mathrm{y}_{\mathrm{ik}}) = \rho(\mu_{\mathrm{ij}}, \mu_{\mathrm{ik}}, \alpha)$$

که ρ یک تابع معلوم است. ضرایب رگرسیونی حاشیه‌ای، β، تفسیری مشابه بـا تفـسیر ضرایب رگرسیونی در آنالیز مقطعی دارند. در مقایسه، مدل‌های حاشیه‌ای برای داده‌های همبسته، مشابه مدل‌های خطی تعمیم‌یافته برای داده‌های مستقل هستند.

هنگامی که رسته‌ها طبیعت رتبه‌ای دارند، مدل‌های لجیت کـه نـوعی از مـدل‌هـای خطی تعمیم‌یافته هستند باید به گونه‌ای باشند که این ویژگی را لحاظ نماینـد. انتخـاب تابع ربط مناسب این امکان را فراهم می‌آورد تا طبیعت رتبـه‌ای بـودن پاسـخ‌هـا، در آن مدل‌بندی در نظر گرفته شود. انـواع توابـع ربـط مناسـب بـرای پاسـخ‌هـای رتبـه‌ای در مدل‌های خطی تعمیم‌یافته را در مکولا[۱۰] می‌توان یافت.

اگرستی[۱۱] مدل‌های لجیت رسته‌های مجاور، لجیـت نـسبت بـه رسـته‌ی آخـر و لجیت تجمعی را برای پاسخ‌های رتبه‌ای به کار برده است.

فرض کنید y پاسخ رسته‌ای مرتب با k رسته باشد. همین طـور $\{\pi_1(x),...,\pi_k(x)\}$

احتمال‌های پاسخ برای مقدار x از مجموعه‌ی متغیرها در نظر گرفته می‌شود. لجیت‌های رسته‌های مجاور عبارتند از:

$$L_k = \log[\pi_k(x)/\pi_{k+1}(x)] \qquad k=1,\dots,K-1$$

لجیت‌های نسبت ادامه به صورت زیر تعریف می‌شوند:

$$L_k = \log\left[\frac{\pi_k(x)}{\pi_{k+1}(x)+\cdots+\pi_K(x)}\right] \qquad k=1,\dots,K-1$$

همچنین لجیت‌های تجمعی را به صورت زیر تعریف می‌کنیم:

$$F_k(x) = \pi_1(x)+\cdots+\pi_k(x) \qquad k=1,\dots,K-1$$

$$L_k = \operatorname{logit}[F_k(x)] = \log\left[\frac{F_k(x)}{1-F_k(x)}\right]$$

$$= \log\left[\frac{\pi_1(x)+\cdots+\pi_K(x)}{\pi_{k+1}(x)+\cdots+\pi_K(x)}\right] \qquad k=1,\dots,K-1$$

ساده‌ترین مدل لجیت تجمعی به صورت زیر است:

$$L_k(x) = \alpha_k \qquad k=1,\dots,K-1$$

که نشان می‌دهد متغیر پاسخ از همه‌ی متغیرهای x مستقل است. α_kها پارامترهای مجزاکننده هستند، برای وارد کردن متغیرها از مدل زیر استفاده می‌کنیم:

$$L_k(x) = \alpha_k + \beta' \qquad k=1,\dots,K-1$$

در این مدل فرض می‌شود اثر متغیرها روی لجیت‌های پاسخ، در رسته‌ی کمتر از k، برای تمام مقادیر k یکسان است. در نتیجه داریم:

$$L_k(x_1) - L_k(x_2) = \log\left[\frac{P(y\le k\mid x_1)/P(y>k\mid x_1)}{P(y\le k\mid x_2)/P(y>k\mid x_2)}\right] = \beta''(x_1 - x_2)$$

به طور کلی مدل‌بندی پاسخ‌های رسته‌ای مکرر، تأثیر متغیرها را بر توزیع‌های حاشیه‌ای پاسخ بررسی می‌کند. برای بیان نحوه‌ی مدل‌بندی فرض کنیم، در زمان t، احتمال پاسخ k وقتی که مجموعه‌ی متغیرها مقدار x را اختیار می‌کند برابر با $\phi_k(t,x)$ باشد و سطوح مختلف x نشان‌دهنده‌ی زیر جامعه‌هایی است که می‌خواهیم توزیع‌های پاسخ آن‌ها را با هم مقایسه نماییم. بنابراین احتمال‌های $\{\phi_1(t,x),...,\phi_k(t,x)\}$، t امین توزیع حاشیه‌ای برای K^t رسته‌ی پاسخ در زیر جامعه‌ی x را تشکیل می‌دهد.

اگر برای همه‌ی پاسخ‌های K در تمامی x ها احتمال $\phi_k(t,x)$ یکسان باشد، در این صورت هیچ تفاوتی بین زیر جامعه‌ها وجود نخواهد داشت. هدف دیگر در این مدل‌بندی، تحلیل نحوه‌ی تغییرات توزیع حاشیه‌ای در زمان‌های $t=1,...,T$ برای یک x ثابت است. در این جا نیز در هر زیر جامعه‌ی x اگر داشته باشیم: $\phi_k(1,x)=\phi_k(2,x)=\cdots=\phi_k(T,x)$، می‌توان گفت هیچ تفاوتی در توزیع‌های حاشیه‌ای در زمان‌های مختلف وجود ندارد. اگر مدل خطی تعمیم‌یافته را به صورت $F(\pi)=X\beta$ نشان دهیم، چون برای توابع ربط لجیت تجمعی داریم[۱۲]:

$$F(\pi)=\beta\log A\pi$$

لذا با انتخاب مناسب ماتریس‌های A و β می‌توانیم توابع ربط لجیت تجمعی را تولید کنیم. در لجیت تجمعی، A تعداد $2(k-1)T$ احتمال حاشیه‌ای تجمعی و مکمل‌های آن را می‌سازد. در هر دو حالت فوق، تبدیل لگاریتم روی تمام عناصر $A\pi$ اعمال شده و هر سطر از ماتریس β که همه‌ی عناصر آن بجز یک مورد ۱ و یک مورد ۱- ، بقیه برابر ه هستند، لجیت مورد نظر را ایجاد می‌کند. برای مثال فرض کنید احتمال تجمعی پاسخ حداکثر k در زمان t برای جامعه‌ی iام را با F_{itk} نشان دهیم. در این صورت:

$$\log\text{it}[F_{itk}] = \log\text{it}[P(y_{it} \le k)] = \log\frac{P(y_{it} \le k)}{1 - P(y_{it} \le k)}$$
$$= \log(\phi_{it1} + \cdots + \phi_{itk}) - \log(\phi_{itk+1} + \cdots + \phi_{itK}) \qquad i = 1,\dots,S \quad t = 1,\dots,T \quad k = 1,\dots,K$$

با توجه به روابط فوق، لجیت تجمعی برابر حاصل‌ضرب ۱ در لگاریتم احتمال تجمعی تا پاسخ k به اضافه‌ی حاصل‌ضرب ۱- در لگاریتم احتمال مکمل از پاسخ k +۱ به بعد است. پس از ماتریس A دو سطر شامل ۰ و ۱ها برای ایجاد این دو احتمال تجمعی و مکمل باید این طور انتخاب شود که در سطر اول، k عنصر اول ۱ و بقیه ۰ باشند و در سطر دوم، k عنصر اول ۰ و بقیه ۱ باشند. به این ترتیب در ماتریس $A\pi$ دو احتمال F_{itk} و $1 - F_{itk}$ به وجود می‌آید. پس از اعمال تبدیل لگاریتم مقادیر $\log F_{itk}$ و $\log(1 - F_{itk})$ ایجاد خواهند شد و در مرحله‌ی بعد کافی است ماتریس β طوری تعیین گردد که عدد ۱ در $\log F_{itk}$ و عدد ۱- در $\log(1 - F_{itk})$ ضرب شود و مجموع این دو حاصل‌ضرب برابر لجیت تجمعی موردنظر خواهد بود.

کاربرد روش در یک مطالعه‌ی واقعی

در یک بررسی روی زنان مبتلا به سرطان پستان، مراجعه‌کننده‌گان به بخش رادیوتراپوتیک انکولوژی بیمارستان امام خمینی (ره) در سال ۱۳۸۱ که از سوزش پوست پس از پرتودرمانی رنج می‌بردند، مورد مطالعه قرار گرفتند. بیماران به دو گروه تصادفی تقسیم شدند. به یک گروه پماد بتامتازون والرات و به گروه دیگر وازلین تجویز شد. شدت واکنش پوستی برای آن‌ها در انتهای یک دوره‌ی درمان ۵ هفته‌ای و سپس ۳ هفته پس از آخرین هفته‌ی درمان، مورد ارزیابی قرار گرفت، یعنی در دو مقطع زمانی، پاسخ بیمار به درد ثبت شد. شدت درد به صورت عدم درد، متوسط، شدید، خیلی شدید می‌باشد و از هر فرد پاسخ شدت درد در دو مقطع زمانی اندازه‌گیری شد. متغیرهای توضیحی از قبیل سن، وضعیت منوپوز، نوع عمل جراحی، سابقه‌ی کموتراپی، نوع درمان و زمان نیز برای بیماران ثبت گردید.

نتایج خروجی‌های برنامه‌ی genmod نشان می‌دهد بجز دو متغیر نوع درمان و زمان، وجود بقیه‌ی متغیرها در مدل ضروری نیست. در مدل‌بندی این داده‌ها در روش حداقل مربعات وزنی با توجه به حجم داده‌ها و مشکلاتی که با زیاد شدن متغیرهای توضیحی و همچنین حضور متغیرهای پیوسته به وجود می‌آید، لازم است متغیرهایی که طبق برنامه‌ی genmod غیر ضروری است از مدل حذف گردد تا براورد پارامترها و تحلیل داده‌ها به روش WLS راحت‌تر انجام شود.

ابتدا از تابع ربط لجیت رسته‌های مجاور استفاده می‌کنیم خروجی به این صورت می‌شود: پی‌مقدار معادل ۰/۸۹ و خی‌دو مربوط به باقی‌مانده برابر ۶/۶۱ می‌باشد، لذا این مدل نسبتاً خوب برازش می‌شود.

سپس از تابع ربط لجیت نسبت به رسته‌ی آخر استفاده می‌کنیم که در این حالت، پی‌مقدار معادل ۰/۸۵ و خی‌دو باقی‌مانده برابر ۶/۹۸ می‌باشد و لذا این مدل نیز نسبتاً خوب برازش می‌شود.

در حالت آخر از تابع ربط لجیت تجمعی استفاده می‌کنیم. در این حالت پی‌مقدار معادل ۰/۹۸ (نزدیک به ۱) و مقدار آماره‌ی خی‌دو باقی‌مانده۴/۲ می‌باشد. بنابراین در مقایسه با مدل‌هایی با لجیت رسته‌های مجاور و لجیت نسبت به رسته‌ی آخر دارای بهترین برازش است.

بحث و نتیجه‌گیری

با توجه به ماهیت داده‌ها در مطالعات طولی با پاسخ رتبه‌ای، در تحلیل این نوع داده‌ها باید تمهیدات خاصی در نظر گرفته شود. در این مطالعات متغیر مورد اندازه‌گیری برای هر آزمودنی بیش از یک بار اندازه‌گیری می‌شود و لذا به علت عدم استقلال پاسخ‌ها در روش اندازه‌گیری مکرر از روش آنالیز رگرسیون معمول استفاده نمی‌گردد، زیرا این روش مستلزم استقلال پاسخ‌هاست. همچنین روش اندازه‌گیری مکرر، بیش از یک واحد مشاهده‌ای را شامل می‌شود و بنابراین پاسخ‌ها به صورت یک سری زمانی ساده نمی‌باشد.

چون متغیر پاسخ، رتبه‌ای است، در انتخاب تابع ربط مناسب باید دقت کرد تـا تـابع ربط انتخابی طبیعت رتبه‌ای بودن پاسخ را لحاظ نماید. در این مطالعـه، ضـمن بررسـی توابع ربط لجیت نسبت به رسته‌ی آخر، لجیت رسته‌های مجاور و لجیت تجمعـی، تـابع ربط لجیت تجمعی نسبت به بقیه برازش بهتری داشت، لذا نتیجه‌ی این تحقیق هماننـد مطالعات قبلی، استفاده از تابع ربط لجیت تجمعی را در مطالعات طـولی بـا پاسـخ‌هـای رتبه‌ای تأیید می‌کند.

مرجع‌ها:

۱) Agresti, A. (1989), A survey of models for repeated ordered categorical response data, *Stat. in Med*., 8, 1209-1224.

۲) Agresti, A., Natarjan, R. (2001), Modeling clustered ordered categorical data, *International Statistical Review*, 69, 345-371.

۳) Pendergast, J.F., Gange, S. J., Newton, M. A., Lindstrom, M. J., Palta, M. and Fisher, M. R. (1996), A Survey of methods for analyzing clustered binary response data, *Internationals Statistical Review*, 64, 89-118.

۴) Zeger, S. L., Liang, K. (1992), An overview of methods for analysis of longitudinal data, *Stat. in Med*., 11, 1825-1839.

۵) Starm, D. O., Wei, L. J. and Ware, J. H. (1988), Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates, *Journal of the American Statistical Association*, 83, 631-637.

۶) Ashby, M., Neuhaus, J. M., Hauck, W. W, Baccetti, P., Meilborn, D. C., Jewell, N. P., Segal, M.R., Fusaro, R. E. (1992), An annotated bibliography of methods for analyzing correlated categorical data, *Stat. in Med*., 11, 67-99.

۷) Kenward, M.G., Lesaffre, E., Molenberghs, G. (1994), An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random, *Biometrics*, 50, 945-953.

۸) علوی مجد، ح.، مقادیر گم‌شده در تحلیل داده‌های رسته‌ای با اندازه‌گیری‌های مکـرر، پایان‌نامه‌ی دکتری آمار زیستی، دانشگاه تربیت مدرس، ۱۳۷۵.

۹) قربانی، راهب، تحلیل پاسخ‌های رسته‌ای رتبه‌ای طولی با مقـادیر گـم‌شـده بـه روش بیزی، پایان‌نامه‌ی دکتری آمار زیستی، دانشگاه تربیت مدرس، ۱۳۸۰.

۱۰) Mcculagh, P.(1980), Regression models for ordinal data, with discussion, *J. R. Statist. Soc*., B, 42, 109-142.

۱۱) Agresti, A. (1990), Categorical data analysis, Wiley, New York.

۱۲) Mccullagh, P. and Nelder, J. A. (1989), Generalized linear models, 2nd ed., London, Chapman and Hall.