

برازش مدل‌های رگرسیونی جمعی و ضربی به نسبت شانس با استفاده از روش منتل - هنزل

دکتر انوشیروان کاظم نژاد^۱، بی بی معصومه میرموسوی^۲، سمانه حسین زاده محمدآباد^۳

یکی از اهداف مهم در تحقیقات اپیدمیولوژی بدست آوردن نسبت‌های شانس می‌باشد. فرض کنید تعداد زیادی (K) جدول 2×2 داشته باشیم که K نسبت به حجم هر یک از جداول N_K بسیار بزرگ باشد، در اینصورت جداول دارای خاصیت *sparse* هستند و همچنین مشاهدات درون هر سطر (ستونها) وابسته و سطرها از یکدیگر مستقل می‌باشند. بنابراین به طور کلی توجه ما معطوف به داده‌های وابسته و *sparse* است. برای جدول k ام، نسبت شانس را با ϕ_k نشان می‌دهیم که ارتباط بین عامل مواجهه و بیماری را توصیف می‌کند. مدل رگرسیونی نسبت شانس به متغیرهای کمکی هر یک از جداول را به صورت $\phi_k = \phi(x(k), \beta)$ نشان می‌دهیم. هدف بدست آوردن برآورد معتبری برای ضرایب رگرسیونی β همچنین آزمون فرض $H_0: \beta = \beta_0$ می‌باشد، زمانی که ساختار وابستگی داده‌ها نامعلوم باشد. همان طور که (*liang*) [۱] نشان داد استنباط ضرایب بر اساس درست‌نمایی فوق هندسی غیر مرکزی به نامشخص بودن ساختار وابستگی داده‌ها حساس است. یک روش مفید دیگر برای برآورد ضرایب رگرسیونی در مدل‌های جمعی و ضربی، روش تابع برآورد منتل - هنزل می‌باشد که یک برآورد سازگار برای پارامترها تولید می‌کند. ما تحت روش تابع برآورد منتل - هنزل برای بدست آوردن فواصل اطمینان از روشهای اسکور، والد و نیمه درست‌نمایی منتل - هنزل استفاده می‌کنیم و سپس با کمک شیبه سازی داده‌ها، به مقایسه روشهای مختلف بدست آوردن فواصل اطمینان می‌پردازیم.

کلمات کلیدی: نسبت شانس، داده‌های وابسته و *sparse*، روش منتل - هنزل،

^۱استادیار گروه آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس
^۲دانشجوی کارشناسی ارشد آمار زیستی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس
^۳دانشجوی کارشناسی ارشد آمار زیستی، دانشکده بهداشت و انستیتو تحقیقاتی، دانشگاه علوم پزشکی تهران

مدلهای رگرسیون جمعی و ضربی ، آماره آزمون اسکور ، والد و نیمه درستنمایی متل هنزل

۱ . مقدمه

یکی از اهداف مهم در تحقیقات اپیدمیولوژی ، برآورد نسبت شانس ϕ ، می باشد که ارتباط بین عامل مواجهه (*exposure*) و بیماری را نشان می دهد. فرض کنید تعداد K جدول 2×2 داشته باشیم که k امین جدول در زیر نشان داده شده است . نسبت شانس در این جدول ، به قرار زیر است :

		مواجهه		
		+	-	
بیماری	+	s_k	$n_k - s_k$	n_k
	-	r_k	$m_k - r_k$	m_k
		t_k	$N_k - t_k$	N_k

$$\phi_k = \frac{s_k \times (m_k - r_k)}{r_k \times (n_k - s_k)} \quad k = 1, \dots, K$$

فرض کنید که این جداول دارای دو ویژگی وابستگی (*dependence*) و *sparsness* باشند .

تعریف *sparsness* : اگر تعداد جداول را با K و حجم جدول k ام را با N_k نشان دهیم بطوریکه k نسبت به N_k بسیار بزرگ باشد آنگاه ویژگی *sparsness* در این جداول برقرار است . به عبارت دیگر وقتی که $K \rightarrow \infty$ ، N_k محدود باقی بماند .

تعریف وابستگی : در این حالت ، در هر جدول مشاهدات مربوط به ستونها به یکدیگر وابسته و سطرها از هم مستقل هستند . معمولاً این حالت در مطالعات طولی یا شاهد - موردی (*case - control*) پیش می آید که در آن شاهد و مورد را براساس یک سری متغیرهای کمکی جور کرده اند و یا پاسخها (مشاهدات) ، مربوط به افراد یک خانواده است .

حال برای هر جدول با خصوصیات ذکر شده ، یک نسبت شانس بدست می آوریم و برای هر نسبت شانس یک معادله رگرسیونی در رابطه با متغیرهای کمکی مربوط به همان جدول می نویسیم . این متغیرهای کمکی می توانند همان متغیرهای مخدوش کننده ای که افراد شاهد - مورد را بر اساس آنها جور کرده ایم باشند . مدل رگرسیونی برای جدول k ام را به صورت $\phi_k = \phi(x_k, \beta)$ نشان می دهیم که در آن x_k یک بردار p بعدی از متغیرهای کمکی و β یک بردار p بعدی از ضرایب رگرسیونی می باشد .

$$x_k = (x_{1k}, x_{2k}, \dots, x_{pk})$$

$$\beta = (\beta_1, \dots, \beta_p)$$

متغیرهای کمکی می توانند گسسته یا پیوسته باشند .

هدف ما ، در این مقاله برآورد ضرایب رگرسیونی β و آزمون فرضیه $H_0 : \beta = \beta_0$ در معادله رگرسیونی می باشد .

۲ . روشهای برآورد ضرایب رگرسیون

جهت برآورد ضرایب β در معادله رگرسیونی $\phi_k = \phi(x_k, \beta)$ دو روش به کار برده می شود :

likelihood approach

(۱) روش درستنمایی

estimating function approach

(۲) روش تابع برآورد (منتل - هنزل)

۱.۲ روش درستنمایی

برسلو (boreslow) [2] مدل رگرسیونی نسبت شانس را برای داده های مستقل شرح داده است و در آن جهت ضرایب رگرسیونی از روش درستنمایی شرطی استفاده نمود و فرض کرد که مشاهدات هر خانه از جدول دارای توزیع دوجمله ای است .
(p_{k1}, p_{k2} ثابت هستند)

$$s_k \sim b(n_k, p_{k1})$$

$$r_k \sim b(m_k, p_{k2})$$

روش درستنمایی شرطی فرض وابستگی داده‌ها را نادیده می‌گیرد و در مقابل توزیع فوق هندسی غیر مرکزی را می‌پذیرد. در این حالت برسلو برای هر جدول توزیع شرطی مشاهدات را با فرض ثابت بودن مجموعهای حاشیه‌ای (t_k, m_k, n_k) در نظر گرفت.

$$p(S_k = s | t_k, m_k, n_k) = \frac{\binom{n_k}{s} \binom{m_k}{t_k - s} \phi_k^s}{\sum_{u=0}^{t_k} \binom{n_k}{u} \binom{m_k}{t_k - u} \phi_k^u}$$

حال با توجه به اینکه $\phi_k(\beta) = \exp(\beta' x_k)$ است داریم:

$$p_k = p(S_k = s | t_k, m_k, n_k) = \frac{\binom{n_k}{s} \binom{m_k}{t_k - s} \exp(s\beta' x_k)}{\sum_{u=0}^{t_k} \binom{n_k}{u} \binom{m_k}{t_k - u} \exp(u\beta' x_k)}$$

$$L(\beta) = \sum_{k=1}^K \log p_k$$

حال با در نظر گرفتن $\frac{\partial L(\beta)}{\partial \beta} = 0$ ، برآورد ضرایب β مشخص می‌شود. از آنجا که جداول مورد بحث ما sparse می‌باشند (تعداد جداول زیاد است) تابع درستنمایی تعداد زیادی پارامتر مزاحم

$$\{\Pi_{\circ k} \text{ کنترل افراد } k = 1, \dots, K\}$$

به وجود می‌آید. به خصوص وقتی $K \rightarrow \infty$ ، تعداد پارامترهای مزاحم افزایش می‌یابد و هر چه این تعداد زیادتر شود برآورد درستنمایی ماکزیمم، برآورد ناسازگارتری خواهد بود. [3]

می‌توان به کمک روشهای خاصی مثل شرطی کردن و یا مدل اثرات تصادفی، پارامترهای مزاحم را حذف نمود که در روش شرطی کردن همانگونه که توسط برسلو [2] نشان داده شد مجبور به پذیرش توزیع دو جمله‌ای برای مشاهدات می‌باشیم که این مطلب برای داده‌های وابسته معتبر نمی‌باشد. لذا نمی‌توان از روش درستنمایی شرطی برای برآورد ضرایب رگرسیون در حالیکه داده‌ها وابسته و sparse هستند، استفاده نمود.

روش دیگری که بتوان از طریق آن ضرایب رگرسیونی را برای داده‌های وابسته و sparse برآورد نمود، روش تابع برآورد می‌باشد.

۲.۲ روش تابع برآورد (منتل - هنزل) :

روش تابع برآورد زمانی استفاده می‌شود که بخواهیم استنباطی راجع به پارامترهای جامعه داشته باشیم ولی توزیع تمام مشاهدات را ندانیم. تابع برآورد تابعی از مشاهدات و پارامترهای مورد نظر است و نتایج بهتری نسبت به روش درست‌نمایی دارد، به این علت که در این روش احتیاج به مشخص بودن و شناختن تنها تعداد محدودی از گشتاور مشاهدات و یا احتمالات می‌باشد ولی در روش درست‌نمایی باید تمام احتمالات موجود باشد. اگر برآورد بردار پارامتر رگرسیونی β که متغیرهای توضیحی را به پارامتر ϕ مرتبط می‌سازد، مورد نظر باشد تابع برآورد β بصورت زیر می‌باشد :

$$h(y, \beta) = \frac{\partial \phi^T}{\partial \beta} q(y, \phi)$$

که $q(y, \phi)$ را تابع مقدماتی گویند [4]. حال می‌توان از روش تابع برآورد به جای روش درست‌نمایی برای برآورد ضرایب رگرسیونی استفاده نمود، که در اینجا به تابع برآوردی که تعمیم برآورد منتل - هنزل است اشاره می‌کنیم.

در مقایسه این روش با روش درست‌نمایی باید اشاره نمود که تابع برآورد منتل - هنزل وقتی که $k \rightarrow \infty$ ، برآورد سازگاری از β می‌دهد، حتی اگر ساختار وابستگی داده‌ها مشخص نباشد.

۱.۲.۲ تابع برآورد منتل - هنزل :

در این روش تابع برآورد مقدماتی به صورت

$$q_k(y_k, \phi_k) = a_k d_k - \phi_k b_k c_k$$

است و مدل رگرسیونی برای نسبت شانس بصورت $\phi_k = \phi(x_k, \beta)$ نشان داده می شود .

تابع برآورد جهت برآورد پارامترهای رگرسیونی به صورت زیر خواهد بود :

$$h(y, \beta) = \sum_{k=1}^K \frac{\partial \phi_k}{\partial \beta} w_k(\phi_k) (a_k d_k - \phi_k b_k c_k)$$

که برای بدست آوردن برآورد پارامتر β باید $h(y, \hat{\beta}) = 0$ قرار دهیم و چون خاصیت $E(h(y, \beta)) = 0$ برقرار است ، بنابراین $\hat{\beta}$ یک برآورد سازگار برای β ، وقتی $K \rightarrow \infty$ می باشد .

وزن $w_k(\phi_k)$ وزنی است که واریانس شرطی مجانبی $\hat{\beta}$ را می نیمم می سازد و عبارتست از :

$$w_k(\phi_k) = \frac{E\{b_k c_k | A_k(\phi_k); \phi_k\}}{\text{var}\{a_k d_k - \phi_k b_k c_k | A_k(\phi_k); \phi_k\}}$$

که شناسه شرطی $A_k(\phi_k)$ برای حذف پارامترهای مزاحم در وزن مورد نظر بکار می رود .

اگر این فرض را که ارتباطی بین عامل مواجهه و بیماری وجود ندارد بپذیریم ، داریم :

$$\begin{aligned} w_k(\phi_k) &= f_k(\phi_k) \frac{E(b_k c_k | A_k; \phi_k = 1)}{\text{var}\{a_k d_k - \phi_k b_k c_k | A_k; \phi_k = 1\}} \\ &= f_k(\phi_k) \left\{ \frac{m_k n_k t_k s_k}{N_k (N_k - 1)} \right\} \left(\frac{m_k n_k t_k s_k}{N_k - 1} \right)^{-1} \\ &= \frac{f_k(\phi_k)}{N_k} \end{aligned}$$

که در اینجا بر روی مجموعه‌های حاشیه‌ای $A_k = (m_k, n_k, t_k, s_k)$ شرط بندی کرده‌ایم .

$w_k(\phi_k)$ شامل یک فاکتور تصحیح می باشد که این فاکتور تصحیح معمولاً برای این انتخاب می شود که $\phi(x_k, \hat{\beta}) = \frac{1}{\phi(x_k, \hat{\beta})}$ برقرار باشد که در آن $\hat{\beta}$ برآورد پارامترهای رگرسیونی است وقتی که جای سطر یا ستون در جدول عوض شود ، یعنی خاصیت تعویض پذیری برقرار شود .

یکی از خواص تابع برآورد منتل-هنزل قابلیت انعطاف پذیری آن در معادلات رگرسیونی است یعنی برای این معادلات، می توان هم از مدل رگرسیونی جمعی و هم ضربی استفاده نمود. مدل رگرسیونی ضربی برای اولین بار توسط لیانگ (Liang) [5] برای داده های مستقل و مدل رگرسیونی جمعی توسط پرنیتیک (Prentic) [6] انجام شد.

وزن $w_k(\phi_k)$ در مدل ضربی برای برقراری خاصیت تعویض پذیری سطر و ستون جدول بکار برده می شود ولی در مدل ضربی معمولاً این وزنها بکار نمی رود. در عمل برای مدل جمعی از وزنهایی مشابه وزنه های مدل ضربی استفاده می شود، به این علت که آزمونهای اسکور و نیمه درستنمایی که جهت آزمون فرض $H_0: \beta = \beta_0$ به کار می روند، در مدل جمعی و ضربی پایا شوند. برای دو مدل ضربی و جمعی، وزنها و تابع برآورد منتل-هنزل در جدول ۱ نشان داده شده است.

	ϕ_k	$w_k(\phi_k)$	تابع برآورد منتل - هنزل: $h(y, \beta)$
مدل جمعی	$1 + x'_k \beta$	$\phi_k^{-\tau/\tau} N_k^{-1}$	$\sum x_k \phi_k^{-\tau/\tau} N_k^{-1} (a_K d_k - \phi_k b_k c_k)$
مدل ضربی	$e^{x'_k \beta}$	$\phi_k^{-\tau/\tau} N_k^{-1}$	$\sum x_k \phi_k^{-1/\tau} N_k^{-1} (a_K d_k - \phi_k b_k c_k)$

جدول ۱

۳. آزمون فرضیه $H_0: \beta = \beta_0$

اغلب تصمیم راجع به اینکه بردار پارامتر بطور معنی داری از مقدار فرض صفر (β_0) متفاوت است یا خیر، هدف اولیه آنالیز به شمار می رود. معمولاً سه آماره آزمون که به صورت مجانبی معادل هستند در دسترس می باشد.

likelihoood ratio

(۱) نسبت درستنمایی

maximum likelihood or wald statistic

(۲) آماره درستنمایی ماکزیمم یا والد

score statistic

(۳) آماره اسکور

mantel - haenszel quasi-likelihood

(۴) تابع درستنمایی منتل - هنزل

از آنجا که از تابع برآورد منتل - هنزل برای برآورد ضرایب رگرسیونی در داده های sparse و وابسته استفاده می کنیم ، دیگر قادر به استفاده از آزمون نسبت درستنمایی نیستیم . اکنون به شرح و بیان سه آزمون دیگر می پردازیم .

۱.۳ آزمونهای والد و اسکور

به طور کلی آماره آزمون والد برای آزمون فرض $H_0: Q\beta = r$ بصورت زیر تعریف می شود :

$$w = [\hat{Q} - r]' [Q \text{var}(\hat{\beta}) Q']^{-1} [Q\hat{\beta} - r]$$

که در آن β بردار p بعدی از پارامترها ، Q یک ماتریس $r \times p$ از اعداد ثابت و r یک بردار ثابت می باشد .

$$\text{var}(\hat{\beta}) = \begin{bmatrix} \sigma_{\beta_1}^2 & \dots & \text{cov}(\beta_1, \beta_k) \\ \sigma_{\beta_k}^2 & \dots & \text{cov}(\beta_k, \beta_1) \end{bmatrix}$$

و دارای توزیع χ^2 (کی دو) با درجه آزادی برابر با تعداد قیدها (تعداد سطرهای ماتریس Q) می باشد

مثلاً اگر فرض $H_0: \beta_1 = \beta_2 = 0$ برقرار باشد و $p = 3$ باشد داریم :

$$Q = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad r = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad w \sim \chi^2(2)$$

حال برای آزمون فرض $H_0: \beta = \beta_0$ آزمونهای والد و اسکور به صورت زیر

هستند :

$$\lambda_w = (\hat{\beta} - \beta_0)' \Omega_{\beta}^{-1} (\hat{\beta} - \beta_0)$$

$$\lambda_s = (h(y, \beta_0))' \Omega_{\beta}^{-1} (\beta_0) h(y, \beta_0)$$

که در آن

$$\Omega_{\beta} = \Omega_{\beta}^{-1}(\hat{\beta}) \Omega_{\beta}(\hat{\beta}) \Omega_{\beta}^{-1}(\hat{\beta})$$

$$\Omega_{\beta} = \sum \frac{\partial \phi'_k}{\partial \beta} \frac{\partial \phi_k}{\partial \beta} w_k^2(\phi_k) (a_k d_k - \phi_k b_k c_k)^2$$

$$\Omega_{\beta} = \sum \frac{\partial \phi'_k}{\partial \beta} \frac{\partial \phi_k}{\partial \beta} w_k^2(\phi_k) (b_k c_k)$$

که تحت فرض H_0 وقتی $k \rightarrow \infty$ و λ_s و λ_w هر دو دارای توزیع χ_p^2 هستند، در موارد خاص وقتی $p = x_k = 1$ ، تابع برآورد منتل - هنزل همانطور که مورد انتظار است برابر است با:

$$\hat{\phi} = \left(\frac{\sum a_k d_k}{N_k} \right) / \frac{\sum (b_k c_k)}{N_k}$$

۲.۳ تابع نیمه درستنمایی منتل - هنزل

تابع نیمه درستنمایی یک روش تقریبی است که از فرض در مورد ساختار وابستگی داده‌ها دوری می‌کند. این تابع انتگرالی از $h(y, \beta)$ به روی ϕ (نه به روی β) است.

$$\gamma(\beta) = \sum_{k=1}^K \int^{\phi(x_k, \beta)} w_k(\phi_k) (a_k d_k - \phi_k b_k c_k) d\phi_k$$

تحت مدل‌های ضربی و جمعی با w_k های یکسان، تابع γ_β یک تابع پایاست. طبق آنچه که گفته شد $w_k(\phi_k) = \phi_k^{-3/2} N_k^{-1}$ ، تحت مدل‌های ضربی و جمعی داریم:

$$\gamma(\beta) = \sum_{k=1}^K \phi_k^{-1/2} N_k^{-1} (a_k d_k - \phi_k b_k c_k)$$

آزمون نسبت نیمه درستنمایی وقتی $k \rightarrow \infty$ ، با آزمونهای والد و اسکور معادل است.

۴. شبیه سازی (Simulation)

برای بررسی استنباطهای انجام شده توسط روشهای مختلف برای داده های وابسته و *sparse* یک مطالعه شبیه سازی توسط نرم افزار *s-plus* انجام شد. برای راحتی کار تنها از یک متغیر کمکی، x_k ، استفاده شد. مدل رگرسیون ضربی بصورت $\phi_k = e^{\beta_0 + \beta_1 x_k}$ با $\beta_0 = 0, \beta_1 = 0, 1/5, 3$ و مدل رگرسیون جمعی

بصورت $\phi_k = 1 + \beta_0 + \beta_1 x_k$ با $\beta_0 = 0, \beta_1 = 0, 0.5, 1.0$ در نظر گرفته شد. متغیر کمکی x_k به صورت گسسته، طوری لحاظ کردیم که از توزیع دو جمله‌ای با $p = 0.5$ پیروی نماید. مشاهدات در هر سطر یا مستقل در نظر گرفته شده‌اند و یا از توزیع بتا-دوجمله‌ای (beta - binomial) با ضریب همبستگی $\rho = 0.3, 0.6$ پیروی کرده‌اند.

فرض شد که π_{0k} ، شیوع بیماری در گروه شاهد، از توزیع بتا با میانگین 0.5 و واریانس 0.125 پیروی کنند.

برای هر یک از حالات فوق یک نمونه 1000 تایی شبیه سازی شد.

برازش داده‌ها به مدل رگرسیونی با استفاده از روش منتل - هنزل انجام شده است و فواصل اطمینان والد، اسکور و نیمه درستنمایی برای این داده‌ها در مدل رگرسیونی ضربی در جدول ۲ و در مدل رگرسیون جمعی در جدول ۳ نشان داده شده است.

با توجه به جدول ۲، فاصله اطمینان هر دو روش نیمه درستنمایی و والد، پوشش احتمالی خوبی داشتند. در مدل رگرسیونی جمعی (جدول ۳)، فاصله اطمینان والد برای پارامترهای β_0, β_1 دارای پوشش احتمالی چوله داری بود، بخصوص وقتی که همبستگی مشاهدات افزایش می‌یافت.

به علاوه، فاصله اطمینان نیمه درستنمایی برای هر دو ضریب β_0, β_1 بهتر از فاصله اطمینانهای والد بود.

تحت هر دو مدل رگرسیون جمعی و ضربی، فاصله اطمینان اسکور پوشش احتمالی ضعیفی را نشان داده است.

ساختار همبستگی	β_1	والد		نیمه راستنمایی اسکور			
		l	u	l	u	l	u
Independent	0.0	2.8	3.1	2.6	3.1	2.8	3.1
		3.1	2.1	3.0	2.0	3.1	2.1
	1.5	3.1	2.4	3.5	1.9	3.1	2.4
		2.5	3.0	2.5	2.6	2.5	3.0
	3.0	2.5	2.1	4.1	1.4	2.6	2.1
		2.3	2.5	2.7	2.1	2.3	2.5
Beta-binomial , $\rho = 0.3$	0.0	3.2	1.8	3.3	1.9	3.2	1.8
		3.6	2.8	3.4	2.4	3.6	2.8
	1.5	3.8	2.3	5.2	2.1	3.8	2.3
		3.6	2.0	5.0	1.7	3.6	2.0
	3.0	2.2	1.3	6.7	0.8	2.2	1.3
		7.3	2.5	10.2	1.9	7.3	2.5
Beta - binomial , $\rho = 0.6$	0.0	3.0	2.5	3.1	2.5	3.0	2.5
		3.2	2.8	3.3	2.7	3.2	2.8
	1.5	3.2	3.4	5.3	2.2	3.3	3.4
		3.8	2.5	5.4	2.0	4.0	2.5
	3.0	1.4	2.5	3.6	1.1	1.6	2.7
		4.5	2.8	8.7	1.8	4.5	3.0

جدول ۲: مدل رگرسیونی ضربی، فاصله اطمینان ۰/۹۵، $K = 150$ ، $N = 1000$ ،
 در سطر اول $m_k = n_k = 5$ و در سطر دوم $m_k = n_k = 20$

ساختار همبستگی	β_1	والد		نیمه راستنمایی اسکور			
		l	u	l	u	l	u
Independent	0	1.8	2.4	2.6	3.1	2.9	2.9
		2.0	2.3	2.2	2.5	2.1	2.4
	5	0.1	5.1	4.3	0.9	2.6	2.3
		2.4	5.0	3.8	2.6	3.6	3.5
Beta-binomial , $\rho = 0.3$	10	0.2	6.3	5.1	0.8	2.4	3.2
		1.9	5.4	4.2	2.3	3.4	3.4
	0.0	1.3	1.0	2.5	2.4	3.0	2.3
		1.6	1.6	2.4	2.5	2.8	2.5
	5	0.1	7.1	7.7	1.5	4.6	3.2
		0.8	5.1	5.6	0.8	3.9	2.3
Beta - binomial , $\rho = 0.6$	10	0.0	7.5	9.4	0.5	4.5	3.0
		0.5	7.6	9.1	0.6	5.9	3.5
	0.0	0.4	0.2	2.8	2.3	2.9	2.4
		0.7	0.5	3.3	2.6	3.3	2.3
	5	0.0	8.8	7.5	0.6	4.1	3.2
		0.1	7.6	6.7	0.7	3.9	3.0
10	0.0	8.4	12.6	0.2	7.1	3.6	
	0.0	8.2	10.4	0.2	6.6	2.6	

جدول ۳: مدل رگرسیونی جمعی، فاصله اطمینان ۰/۹۵، $K = 150$ ، $N = 1000$ ،
در سطر اول $m_k = n_k = 5$ و در سطر دوم $m_k = n_k = 20$

۵. نتیجه گیری

وقتی داده‌ها وابسته و *sparse* هستند از روش رگرسیون لجستیک شرطی نمی‌توان استفاده کرد زیرا همانطور که لیانگ (Liang) [7] نشان داد این روش برآورد آریبی برای ضرایب رگرسیونی در نسبت شانس تولید می‌کند، حتی اگر این همبستگی مقدار متوسط $\rho = 0/3$ باشد.

به علاوه لیانگ نشان داد که روش منتل-هنزل در مورد داده‌های مستقل با روش درستنمایی شرطی معادل است و استنباط بهتری برای داده‌های وابسته نسبت به روش درستنمایی شرطی می‌دهد.

تحت روش منتل-هنزل برای بدست آوردن فواصل اطمینان پارامترهای رگرسیون از سه آزمون والد، اسکور و نیمه درستنمایی استفاده نمودیم. با توجه به نتایج جداول ۲ و ۳، فاصله اطمینان والد در مدل رگرسیون ضربی رفتار خوبی نشان داده است که در مورد رگرسیون جمعی اینگونه نبوده است. فاصله اطمینان اسکور نیز تحت هر دو مدل ضربی و جمعی رفتار ضعیفی نشان داده است ولی فاصله اطمینان نیمه درستنمایی منتل-هنزل تحت هر دو مدل خوش رفتار بوده است.

بنابراین روش نیمه درستنمایی منتل-هنزل از اعتبار بهتری نسبت به دو روش اسکور و والد برخوردار است، گرچه تنها استثنا در این اینجا زمانی است که در مدل رگرسیون جمعی، β_1 و ضریب همبستگی ρ بزرگ باشند و در این حالت شبیه سازی و نتایج حاصل از آن نشان داد که هیچ یک از روشهای فوق بطور کامل مناسب نیستند.

۶. مراجع

- [1] - Liang , K.-Y. (1985) . Odds ratio inference with dependent data . *Biometrika* 72 , 678-682 .
- [2] - Breslow , N . (1976) . Regression analysis of the log odds ratio : A method for retrospective studies . *Biometrics* 32 , 409 - 416 .
- [3] - Kalbfleisch , J. D. and Sprott , D. A. (1970) . Application of likelihood methods to models involving large numbers of nuisance parameters (with discussion) . *Journal of the Royal Statistical Society , Series B* 32 , 175 - 208 .
- [4] - Hanfelt, J. J. and Liang , K. Y. (1995) . Approximate likelihood ratios for general estimating functions. *Biometrika* 82 , 461 - 477 .
- [5] - Liang , K. - Y. , Beaty , T. H. , and Cohen , B. H. (1986) . Application of odds ratio regression models for assessing familial aggregation from case - control studies. *American Journal of Epidemiology* 124 , 678 - 683 .

[6] - Prentic , R. L. and Mason , M. W. (1986) . On the application of linear relative risk regression models. *Biometrics* 42 , 109 - 120 .

[7] - Hanfelt , J. J. and Liang , K. - Y. (1998) . Inference for odds ratio regression models with sparse dependent data . *Biometrics* 54 , 136 - 147