مجله‌ی بررسی‌های آمار رسمی ایران
سال ۳۰، شماره‌ی ۱، بهار و تابستان ۱۳۹۸، صص ۵۵– ۶۶

تعیین بیزی اندازه نمونه در مطالعات طولی با پاسخ‌های دودویی با استفاده از مدل اثرهای تصادفی

مریم سوری و مجتبی گنجعلی*

دانشگاه شهید بهشتی

چکیده: تعیین اندازه‌ی نمونه در تمام مطالعات آماری از جمله مطالعه‌های طولی، از اهمیت به‌سزایی برخوردار است. این مهم، معمولاً با در نظر گرفتن یک توان معلوم در آزمون فرض‌های آماری برای کاهش هزینه‌های نمونه‌گیری انجام می‌شود. انتخاب اندازه‌ی نمونه‌ی مناسب با استفاده از روش‌های کارآمد، به محقق این اطمینان را می‌دهد که با صرف حداقل انرژی، وقت و هزینه به هدف مطالعه با توان بالا دست یابد. در این مقاله، با استفاده از روشی بر مبنای شبیه‌سازی و معیار توان بیزی، به تعیین تعداد اندازه‌ی مورد نیاز در مطالعات طولی با پاسخ‌های دودویی از دیدگاه بیزی می‌پردازیم. نهایتاً، با ذکر یک مثال کاربردی روش پیشنهادی به کار گرفته شده است.

واژگان کلیدی: تحلیل بیزی؛ مطالعه‌ی طولی؛ مدل با اثرهای تصادفی؛ معیار توان بیزی.

۱- مقدمه

تعیین اندازه‌ی نمونه در پژوهش‌های آماری از آنجا که بر مبنای تحلیل نمونه‌ی کافی نشان‌دهنده‌ی واقعیت‌های جامعه است، نقشی اساسی ایفا می‌کند. هرچه تعداد یا اندازه نمونه بزرگ‌تر باشد، میزان خطا در نتیجه‌گیری کاهش می‌یابد اما از طرفی افزایش اندازه‌ی نمونه مستلزم صرف زمان، انرژی و هزینه‌ی بالا است، بنا بر این روش‌های آماری تعیین اندازه‌ی نمونه‌ی کارآمد، با کم‌ترین خطا، از اهمیت به سزایی برخوردارند.
تاکنون روش‌های آماری برای تحلیل داده‌ها در آزمایش‌های بالینی تصادفی و یا مطالعات مشاهده‌ای برای پاسخ‌های همبسته، به‌خوبی توسعه یافته‌اند. همچنین روش‌های مناسب

* نویسنده‌ی عهده‌دار مکاتبات
دریافت: ۱۳۹۹/۳/۱، پذیرش: ۱۳۹۹/۸/۱۵.

اندازه‌گیری اندازه‌ی نمونه برای برنامه‌ریزی چنین مطالعاتی، در مطالعات مقطعی (طی سال‌های ۱۹۸۱-۲۰۱۰) مورد توجه محققان قرار گرفته است. به‌عنوان مثال [۷] و [۹] محاسبه‌ی توان برای مدل‌های چندمتغیره‌ی خطی عام را از طریق روش‌های آزمایش تک‌متغیری مورد بررسی قرار دادند. در این میان، علاوه بر مطالعات مقطعی، مطالعه‌های طولی از اهمیت به‌سزایی برخوردارند. مطالعه‌ی طولی، مطالعه‌ای است که در آن متغیر پاسخ برای آزمودنی‌ها چندین بار در طول زمان اندازه‌گیری می‌شود. به دلیل اندازه‌گیری‌های مکرر، در این دسته از مطالعات امکان مطالعه‌ی تغییرات درون هر واحد وجود دارد. بنابراین مطالعات طولی در علوم اجتماعی و علوم پزشکی مورد توجه محققان قرار گرفته است. در طرح‌های طولی، اثر مداخله‌گر با اثر کنترل، توسط معنی‌داری اثر متقابل زمان و تیمار، مقایسه می‌شوند. علاوه بر این، مدل‌های با اثرات آمیخته به طور معمول برای همبستگی پاسخ‌های با اندازه‌گیری‌های مکرر مورد استفاده قرار می‌گیرند. در این مدل‌ها اثر متقابل زمان و تیمار ثابت در نظر گرفته می‌شوند و اثرات آزمودنی (به‌عنوان مثال، رهگیری و روند زمان) به صورت تصادفی فرض شده‌اند. مادامی که اثرات آزمودنی به صورت تصادفی در نظر گرفته می‌شوند، مدل‌های با اثرات آمیخته نه تنها اطلاعات در مورد متوسط روند ارائه می‌کنند، بلکه برآوردی از تأثیر فرد– ویژه را ارائه می‌دهند. [۸]، [۱۰] و [۴] روش‌هایی را از دیدگاه فراوانی‌گرا برای تعیین اندازه‌ی نمونه در مطالعات طولی، بر مبنای تحلیل توانی ارائه کرده‌اند که اولین گام در محاسبه‌ی اندازه‌ی نمونه‌ی مورد نیاز است. مروری بر این روش‌ها توسط [۱] و [۲] آورده شده است. در عمل مدل‌های آمیخته متداول‌ترین روش برای تحلیل داده‌های طولی است. با این وجود، روش‌های مبنی بر تحلیل توان برای مدل‌های آمیخته اغلب با تقریب تابع درستنمایی همراه است. به‌عنوان مثال، [۵]، روش‌هایی را بر مبنای تقریب تابع درستنمایی در مدل‌های با اثرات آمیخته تصادفی برای رسیدن به تعداد نمونه‌ی مورد نیاز در طرح‌های طولی با پاسخ‌های دودویی، ارائه دادند. با این حال، روش‌های بیزی تعیین اندازه‌ی نمونه نیز مورد توجه محققان قرار گرفته است. به‌عنوان مثال روشی بر مبنای شبیه‌سازی توسط [۱۲]، برای محاسبه‌ی اندازه‌ی نمونه ارائه شده است. مروری بر روش‌های بیزی تعیین اندازه‌ی نمونه را می‌توان در [۱۱] و [۶] یافت.

در این مقاله، ابتدا به معرفی مدل اثرات تصادفی در تحلیل بیزی داده‌های طولی و چگونگی استخراج نمونه از توزیع پسینی آن می‌پردازیم. سپس با استفاده از رویکرد [۳]

در تعیین اندازه‌ی نمونه در مطالعات طولی با پاسخ‌های پیوسته، به دنبال رویکرد بیزی تعیین اندازه‌ی نمونه‌ی مورد نیاز در مطالعات طولی با پاسخ‌های دودویی هستیم. به این منظور، ابتدا مدل‌های اثرات تصادفی را معرفی کرده، سپس با استفاده از روش مبتنی بر شبیه‌سازی [۱۲] و با تعریف یک متغیر پنهان، تعداد اندازه‌ی نمونه مورد نیاز را از دیدگاه بیزی در مدل‌های طولی با اثرات آمیخته، برای پاسخ‌های دودویی به‌دست می‌آوریم.

۲- مدل‌بندی داده‌های طولی با مدل اثرات تصادفی از دیدگاه بیزی

مدل با اثرهای تصادفی روشی برای تحلیل مطالعات طولی است. در این مدل فرض همبستگی میان پاسخ‌های تکرار شده به وسیله‌ی ضرایب رگرسیونی نشان داده می‌شوند. در مدل با اثرهای تصادفی، پاسخ به‌عنوان تابعی از متغیرهای تبیینی با ضرایب رگرسیونی‌ای فرض شده، که از یک آزمودنی به آزمودنی دیگر متفاوت است. این تفاوت‌ها اغلب به خاطر عوامل اندازه‌گیری نشده‌ای است که منتج از عوامل طبیعی و ژنتیکی است. مثلاً رگرسیون خطی ساده برای رشد کودک را در نظر گیریم که ضرایب مدل، بیانگر وزن در بدو تولد و شاخص رشد است. واضح است که کودکان در بدو تولد، به دلیل اختلاف در شرایط محیطی و ژنتیکی، دارای وزن و شاخص رشد متفاوت‌اند. از این رو برای این مثال، یک مدل ضرایب تصادفی مناسب است، زیرا 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.

مدل با اثرهای آمیخته‌ی

$$Y_{it} = \boldsymbol{X}'_{it}\boldsymbol{\beta} + \boldsymbol{z}'_{it}\boldsymbol{\upsilon}_i + \epsilon_{it}\,, \quad i = ۱, \dots, n, \qquad t = ۱, \dots, T \tag{۱}$$

را در نظر گیریم، که در آن Y_{it} پاسخ مربوط به iامین فرد در زمان tام و $\boldsymbol{X}_{it} = (x_{it۱}, \dots, x_{itp})'$ بردار شامل متغیرهای کمکی مربوط به iامین آزمودنی در زمان t، و $\boldsymbol{z}_{it}$ بردار شامل عناصر $(۱, t_i)'$ است. برای هرi، $\boldsymbol{\upsilon}_i = (\upsilon_{۰i}, \upsilon_{۱i})$ را بردار شامل اثرات تصادفی در نظر می‌گیریم که دارای توزیع نرمال با میانگین $\underline{۰}$ و ماتریس واریانس‌کوواریانس $\mathbf{D}$ است. در این مدل $\upsilon_{۰i}$، عرض از مبدأ تصادفی و $\upsilon_{۱i}$، شیب تصادفی برای فرد iام است. همچنین خطاهای ϵ_{it} از هم مستقل بوده و دارای توزیع نرمال با میانگین صفر و واریانس $\sigma^۲$ هستند.

فرم گسترده‌ی مدل (۱) به صورت

$$Y_{it} = \boldsymbol{X}'_{it}\boldsymbol{\beta} + v_{\circ i} + v_{\circ i}t_i + \epsilon_{it}\,, \quad i = ۱, \dots, n, \qquad t = ۱, \dots, T$$

نوشته می‌شود. برای انجام استنباط بیزی، در نظر گرفتن توزیع‌های پیشینی برای پارامترهای نامعلوم ضروری است. پارامترهای نامعلوم در مدل (۱) عبارت‌اند از $\boldsymbol{\beta} = (\beta_\circ, \beta_۱, \dots, \beta_p)'$، $\sigma^۲$ و D توزیع‌های پیشینی برای این پارامترهای نامعلوم می‌تواند به صورت

$$\begin{aligned} \boldsymbol{\beta} &\sim N_p(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \Psi_{\boldsymbol{\beta}}) \\ \sigma^۲ &\sim I\Gamma(a, b) \\ \boldsymbol{D} &\sim IWishart\ (\boldsymbol{R}, v) \end{aligned} \qquad (۲)$$

در نظر گرفته شوند. که در آن، $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ و $\boldsymbol{\Psi}_{\boldsymbol{\beta}}$ به ترتیب میانگین و ماتریس واریانس‌کوواریانس ابرپارامتر $\boldsymbol{\beta}$ هستند. همچنین $\mathrm{I}\Gamma(\mathrm{a}, \mathrm{b})$ توزیع گامای وارون با پارامترهای a و b و $\mathrm{IWishart}(\boldsymbol{R}, v)$ توزیع ویشارت وارون با پارامترهای $\boldsymbol{R}$ و v است.

توزیع پسینی توأم برای پارامترهای نامعلوم $\boldsymbol{\beta} = (\beta_\circ, \beta_۱, \dots, \beta_p)'$، $\sigma^۲$ و D و اثرهای تصادفی $\boldsymbol{v}_i = (v_{\circ i}, v_{۱i})$ ، برای $i = ۱, \dots, n$ عبارت‌اند از:

$$\begin{aligned} \pi(\boldsymbol{\beta}, \sigma^۲, \boldsymbol{D}, \boldsymbol{v}_۱, \dots, \boldsymbol{v_n} | \boldsymbol{y}, \boldsymbol{x}) \propto \textstyle\prod_{i=۱}^{n} \prod_{t=۱}^{m} \Phi(y_{it}; \boldsymbol{x}'_{it}\boldsymbol{\beta} + v_{\circ i} + v_{۱i}t_t, \sigma^۲) \\ \times \emptyset_p(\boldsymbol{\beta}; \boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Psi}_{\boldsymbol{\beta}}) \times \pi\left(\sigma^۲\right) \\ \textstyle\prod_{i=۱}^{n} \emptyset_q(\boldsymbol{v}_i, \circ, \boldsymbol{D}) \\ \times \pi(\boldsymbol{D}; \boldsymbol{R}, v) \end{aligned} \qquad (۳)$$

برای نمونه‌گیری از توزیع پسین توأم (۳)، روش نمونه‌گیری گیبز مورد استفاده قرار می‌گیرد. برای این منظور، توزیع شرطی کامل $\boldsymbol{\beta}$، $\sigma^۲$، $\boldsymbol{D}$ و $\boldsymbol{v}_i$، برای $i = ۱, \dots, n$ مورد نیاز است که عبارت‌اند از:

$$\boldsymbol{\beta} | \boldsymbol{y}, \boldsymbol{x}, \sigma^۲, \boldsymbol{D}, \boldsymbol{v}_۱, \dots, \boldsymbol{v}_n \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}|\dots}, \textstyle\sum_{\boldsymbol{\beta}|\dots})$$

که در آن

$$\boldsymbol{\mu}_{\boldsymbol{\beta}|\dots}(\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-۱}\boldsymbol{\mu}_{\boldsymbol{\beta}} + \frac{۱}{\sigma^۲}\textstyle\sum_{i,t} \boldsymbol{x}_{it}(y_{it} - \boldsymbol{z}'_{it}\boldsymbol{v}_i)' \textstyle\sum_{\boldsymbol{\beta}|\dots})$$

$$\Sigma_{\boldsymbol{\beta}|\ldots} = \left(\sum_{i,t} \frac{\boldsymbol{X}_{it}\boldsymbol{X}'_{it}}{\sigma^{۲}} + \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-۱}\right)^{-۱}$$

و همچنین

$$\sigma^{۲} \Big| \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{v}_{۱}, \ldots, \boldsymbol{v}_n \sim \Gamma\Gamma\left(\frac{nm}{۲} + a, \frac{۱}{۲}\sum_{i,t}\left((y_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta} - \boldsymbol{z}'_{it}\boldsymbol{v}_i)^{۲} + b\right)\right.$$

$$\boldsymbol{D}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{v}_{۱}, \ldots, \boldsymbol{v}_n \sim IWishart(N + V, \textstyle\sum_i \boldsymbol{v}_i\boldsymbol{v}'_i + R)$$

$$\boldsymbol{v}_i|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\beta}, \sigma^{۲}, \boldsymbol{D} \sim N(\boldsymbol{\mu}_{\boldsymbol{v}_i|\ldots}, \textstyle\sum_{\boldsymbol{v}_i|\ldots}) \qquad (۴)$$

که در آن $\boldsymbol{\mu}_{\boldsymbol{v}_i|\ldots} = \sum_t \boldsymbol{z}_{it}\,(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta}) \times \Sigma_{\boldsymbol{v}_i|\ldots}^{-۱}$ و $\Sigma_{\boldsymbol{v}_i|\ldots} = (\sum_t \frac{\boldsymbol{z}_{it}\boldsymbol{z}'_{it}}{\sigma^{۲}} + \boldsymbol{D})^{-۱}$

۳- تعیین اندازه‌ی نمونه به روش بیزی در مطالعات طولی

مبتنی بر شبیه‌سازی [۱۲] به تعیین اندازه‌ی نمونه از دیدگاه بیزی می‌پردازیم. این روش بر مبنای تعریف دو مجموعه از توزیع‌های پیشینی است. توزیع‌های پیشینی مجموعه‌ی اول برای مدل (۱) به منظور تحلیل داده‌ها استفاده می‌شود و ساختار آن‌ها مشابه با توزیع‌های پیشینی در سایر شاخه‌های آمار است. این پیشینی‌ها، پیشین‌های «تحلیل» یا «برازش» نامیده می‌شوند. در روش تعیین اندازه‌ی نمونه با کمک معنی‌داری ضریب رگرسیونی به روش فراوانی‌گرا، پارامترهای نامعلوم با استفاده از برآوردهای طراحی شده با کمک یک متخصص، بر مبنای یک مجموعه داده اولیه یا عملکردی محافظه کارانه انتخاب می‌شود. [۱۲] در انتخاب متغیر بیزی به جای استفاده از مقادیر ثابت پارامترها، استفاده از توزیع‌های پیشینی احتمالی استخراجی را پیشنهاد دادند که به عدم قطعیت در این مقادیر اشاره دارد و مجموعه‌ی دوم پیشینی‌های مورد استفاده است. این توزیع‌های پیشین را پیشین‌های «نمونه‌گیری» یا «طرح» می‌نامند. در روش تعیین اندازه‌ی نمونه از دیدگاه شبیه‌سازی [۱۲]، ابتدا پارامترها را با استفاده از پیشین طرح شبیه‌سازی کرده و سپس با استفاده از پارامترهای تولید شده از پیشین طرح، یک نمونه‌ی تصادفی از مدل موردنظر تولید می‌کنند و سپس برآورد پارامترها را با استفاده از پیشین‌های برازش برآورد می‌کنند که برخلاف پیشین‌های طرح که نسبتاً با اطلاع هستند، معمولاً این پیشین‌ها بی‌اطلاع یا تخت در نظر گرفته می‌شوند. در روش تعیین اندازه‌ی نمونه با کمک معنی‌داری ضرایب

رگرسیونی، اندازه‌ی نمونه‌ی مورد نیاز برای دستیابی به تفاوت قابل قبول معنی‌داری دوری یک ضریب رگرسیونی از صفر، به‌عنوان اندازه‌ی نمونه‌ی بهینه در نظر گرفته می‌شود.
فرض می‌کنیم که $\boldsymbol{D}^{(n)} = \left(\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)}\right)$، مجموعه‌ای از داده‌ها با اندازه‌ی n است که با استفاده از پیشین طرح تولید شده‌اند. پیشگوی خطی برای مدل (۱) را برای $i = ۱, \dots, n$ و $t = ۱, \dots, T$ به فرم

$$\mu_{it} = \beta_{\circ} + \beta_{۱} X_{i۱} + \sum_{k=۲}^{p} \beta_i X_{ik} + v_{\circ i} + v_{۱i} t_t$$

در نظر می‌گیریم که در آن $X_{۱}$ متغیر کمکی نشانگر گروه کنترل یا تیمار مورد نظر است، بنابراین $\beta_{۱}$ ضریب رگرسیونی مطلوب برای تعیین اندازه‌ی نمونه است. به‌منظور تعیین اندازه‌ی نمونه به روش بیزی، از معیارهای مختلفی مانند معیار توان بیزی (BPC)[۱] استفاده می‌شود. در ادامه به تعریف این معیار می‌پردازیم.

معیار توان بیزی: اگر رابطه‌ی بین متغیر کمکی و متغیر پاسخ مثبت باشد، BPC بر مبنای اندازه‌ی نمونه‌ی n برای احتمال‌های α و $۱ - \beta$، عبارت است از:

$$E\left[I\left(P\left(\beta_{۱} > \circ \middle| \boldsymbol{D}^{(n)}\right)\right) > ۱ - \alpha\right] \geq ۱ - \beta \qquad (۵)$$

که در آن $I(.)$ یک تابع نشانگر است. انتخاب‌های رایج برای α مقادیر ۰٫۰۱، ۰٫۰۵ یا ۰٫۱ است. علاوه بر آن $۱ - \beta$ معمولاً مقادیر ۰٫۸ یا ۰٫۹ انتخاب می‌شود. در این مقاله، با استفاده از معیار توان بیزی به تعیین اندازه‌ی نمونه در مطالعات طولی با پاسخ‌های دودویی می‌پردازیم. در واقع برای هر مجموعه داده $\boldsymbol{D}^{(n)}$ فرض $H_{\circ}: \beta_{۱} < \circ$ در مقابل $H_{۱}: \beta_{۱} > \circ$ رد می‌شود.

اگر $I\left(P\left(\beta_{۱} > \circ \middle| \boldsymbol{D}^{(n)}\right)\right) > ۱ - \alpha$، با استفاده از معادله (۵)، دنبال اندازه‌ی نمونه‌ای هستیم که چنین آزمون‌هایی حداقل $۱۰۰\left(۱ - \beta\right)\%$ از دفعات رد شده است. اگر رابطه‌ی بین متغیر پاسخ با متغیر کمکی منفی باشد، آنگاه باید توان را به صورت $E\left[I\left(P\left(\beta_{۱} < \circ \middle| \boldsymbol{D}^{(n)}\right)\right) > ۱ - \alpha\right] \geq ۱ - \beta$ تعریف کرد.

به طور کلی می‌توان الگوریتم شبیه‌سازی تعیین اندازه‌ی نمونه از دیدگاه بیزی را برای داده‌های طولی به صورت زیر در نظر گرفت:

۱- مشخص کردن اثر پارامتر مورد علاقه β_1، و مشخص کردن پیشین‌های طرح و پیشین‌های تحلیل.

۲- در هر اندازه‌ی نمونه‌ی n مراحل زیر M بار تکرار می‌شوند:

الف- تولید مقادیر پارامترهای رگرسیونی، واریانس و پارامتر ضریب همبستگی با استفاده از توزیع‌های پیشینی طرح

ب- شبیه‌سازی متغیرهای کمکی پیوسته از یک توزیع پیوسته، از یک دامنه مناسب و شبیه‌سازی متغیرهای کمکی دودویی از توزیع‌های دودویی با احتمال موفقیت p.

ج- تولید $\boldsymbol{D}^{(n)} = \left(\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)}\right)$ با توجه به فروض توزیعی مربوطه

د- تحلیل داده‌های تولید شده در مرحله‌ی ج، با استفاده از پیشین‌های تحلیل

ه- محاسبه‌ی معیار BPC برای ضریب رگرسیونی مورد علاقه.

۳- برازش یک منحنی برای هر یک از معیارهای BPC با استفاده از اندازه نمونه‌های متفاوت.

۳-۱- تعیین اندازه‌ی نمونه به روش بیزی در مطالعات طولی با پاسخ‌های دودویی

در این بخش، با استفاده از روش پیشنهادی [۱۲] که در بخش قبل معرفی شد، به دنبال رویکرد بیزی تعیین تعداد اندازه‌ی نمونه‌ی مورد نیاز در مطالعات طولی با پاسخ‌های دودویی خواهیم بود. در ادامه یک مثال کاربردی در این زمینه ارائه و نتایج مورد بررسی قرار خواهد گرفت.

پیشگوی خطی مدل (۱) را برای $i = 1, \dots, n$ و $t = 1, \dots, T$، به صورت

$$\mu_{it} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + v_{0i} + v_{1i} t_t \qquad (۶)$$

در نظر گیریم که در آن X_1 متغیر کمکی نشانگر گروه کنترل یا تیمار مورد نظر است. فرض کنید متغیر پنهان Y_{it}^* به صورت

$$Y_{it}^* = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + v_{0i} + v_{1i} t_t + \varepsilon_{it} \qquad (۷)$$

است. اگر قرار دهیم

$$Y_{it} = \begin{cases} ۱ & Y_{it}^* > ۰ \\ ۰ & ow \end{cases}$$

آن‌گاه:

$$P\left(Y_{it} = ۱\right) = P(\, Y_{it}^* > ۰\,)$$

در نتیجه طبق (۷) داریم:

$$\begin{aligned} P\left(Y_{it} = ۱\right) &= P(\varepsilon_{it} > -\mu_{it}) \\ &= P(\varepsilon_{it} < \mu_{it}) \\ &= F_{\varepsilon_{it}}(\mu_{it}) \\ &= \frac{e^{\mu_{it}}}{۱ - e^{\mu_{it}}} \end{aligned} \tag{۸}$$

که در آن $F_{\varepsilon_{it}}(\mu_{it})$ توزیع لوژستیک استاندارد است. در این صورت جفت نمونه‌های $\boldsymbol{D}^{(\boldsymbol{n})} = \left(\boldsymbol{X}^{(\boldsymbol{n})}, \boldsymbol{y}^{(\boldsymbol{n})}\right)$، که در آن $\boldsymbol{y}^{(\boldsymbol{n})}$ برداری از پاسخ‌های دودویی است، با توجه به فروض توزیعی و پیشگوی (۶) تولید می‌شوند. از طرفی با استفاده از رابطه‌ی (۸)، توزیع پسینی در این حالت به صورت

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{v}_۱, \dots, \boldsymbol{v}_n | \boldsymbol{y}, \boldsymbol{x}) \propto & \prod_{i=۱}^{n} \prod_{t=۱}^{m} \left(\frac{e^{\mu_{it}}}{۱+e^{\mu_{it}}}\right)^{y_{it}} \left(\frac{۱}{۱+e^{\mu_{it}}}\right)^{۱-y_{it}} \\ & \times \emptyset_p(\boldsymbol{\beta}; \boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Psi}_{\boldsymbol{\beta}}) \\ & \prod_{i=۱}^{n} \emptyset_q(\boldsymbol{v}_i, ۰, \boldsymbol{D}) \\ & \times \pi(\boldsymbol{D}; \boldsymbol{R}, v) \end{aligned} \tag{۹}$$

است که با استفاده از روش نمونه‌گیر گیبز، از توزیع پسینی (۹) نمونه استخراج می‌شود.

۴ – مثال کاربردی

متغیر پیوسته Y با $T = ۳۵$ اندازه‌ی تکرار شده را در نظر گیرید که برای n آزمودنی ثبت شده است. متغیر نشانگر $X_۱$ را به‌عنوان نشانگر گروه تیمار در نظر بگیرید که $X_۱ = ۱$ به گروه تیمار و $X_۱ = ۰$ به گروه کنترل اشاره دارد. همچنین فرض می‌کنیم که متغیر کمکی

پیوسته‌ی دیگری که آن را $X_۲$ می‌نامیم وجود دارد. بنابراین در این مدل برای $i = ۱, \dots, n$ و $t = ۱, \dots, T$، پیشگوی خطی عبارت است از

$$\mu_{it} = \beta_۰ + \beta_۱ X_{i۱} + \beta_۲ X_{i۲} + v_{۰i} + v_{۱i} t_t$$

که در آن، بردار $\boldsymbol{v}_i = \left(v_{۰i}, v_{۱i}\right)'$ دارای توزیع $N_۲\left(۰, \boldsymbol{D}\right)$ است. متغیر کمکی $X_۱$ به تعداد مساوی به گروه تیمار و کنترل اختصاص داده شده و از توزیع نرمال استاندارد تولید شده است. توجه داریم که هدف ما، تعیین اندازه‌ی نمونه بر مبنای $\beta_۱$ است. بنابراین فرض می‌کنیم در این مثال، علاقمند به تشخیص اثر $\beta_۱$ برابر ۲ هستیم لذا $\beta_۱ = ۲$ را ثابت در نظر می‌گیریم (برای اطلاعات بیشتر به [۱۲]، مراجعه کنید). پیشین‌های طرح و تحلیل برای این مثال در جدول ۱ نشان داده شده‌اند.

جدول ۱- توزیع‌های پیشین طرح و تحلیل، در تعیین اندازه‌ی نمونه در مطالعات طولی با پاسخ دودویی

پارامترها	پیشین طرح	پیشین تحلیل
$\beta_۰$	$N(-۱, ۰/۲)$	$N(۰, ۱۰۰۰)$
$\beta_۱$	تباهیده در ۲	$N(۰, ۱۰۰۰)$
$\beta_۲$	$N(۲, ۰/۲۵)$	$N(۰, ۱۰۰۰)$
$\sigma^۲$	$U(۱۰, ۱۰۰)$	$I\Gamma(۰/۰۰۱, ۰/۰۰۱)$
$\boldsymbol{D}$	$Wishart(\boldsymbol{I}_۲, ۲)$	$IWishart(\boldsymbol{R}, ۲)^*$

$^*\boldsymbol{R} = diag(۰/۰۱, ۰/۰۱)$

طبق روش پیشنهادی در بخش قبل، با تعریف متغیر پنهان Y_{it}^* جفت نمونه‌های $\boldsymbol{D}^{(n)} = \left(\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)}\right)$، که در آن $\boldsymbol{y}^{(n)}$ برداری از پاسخ‌های دودویی است، با توجه به فروض توزیعی و پیشگوی (۶) تولید می‌شوند. جدول ۲، اندازه‌های نمونه‌ی مورد نیاز را برای توان‌های بیزی مختلف شامل ۰/۸ و ۰/۹۵ و تعداد تکرارهای T = ۳/۵ نشان می‌دهد. همانطور که مشاهده می‌شود اندازه‌ی نمونه‌ی مورد نیاز با افزایش توان بیزی، افزایش چشم‌گیری پیدا می‌کند. همچنین همان‌طور که انتظار می‌رود، این جدول کاهش تعداد اندازه‌ی نمونه را برای رسیدن به سطح خاصی از توان، با افزایش تعداد اندازه‌های مکرر نشان می‌دهد.

۵- بحث و نتیجه‌گیری

در این مقاله، به تعیین اندازه‌ی نمونه از دیدگاه بیزی در مطالعات طولی با پاسخ‌های دودویی پرداخته شده است. با استفاده از روش مبتنی بر شبیه‌سازی و به‌کارگیری عامل توان بیزی، مشاهده کردیم که با افزایش توان بیزی، تعداد اندازه‌ی نمونه‌ی مورد نیاز برای انجام آزمایش افزایش می‌یابد. همچنین همان‌گونه که انتظار می‌رود، با افزایش تعداد اندازه‌های مکرر، تعداد نمونه‌ی مورد نیاز برای رسیدن به سطح خاصی از توان بیزی کاهش می‌یابد. لازم به ذکر است در این مقاله، از آنجا که با تعریف متغیر پنهان در مقایسه با [۳] با فقدان اطلاعات مواجه‌ایم و تنها مثبت یا منفی بودن متغیر پنهان را مشاهده می‌کنیم، به تعداد نمونه‌ی بیشتری جهت رسیدن به توان یکسان نیازمندیم.
در این مقاله، با فرض نرمال بودن اثرات تصادفی به تعیین اندازه‌ی نمونه از دیدگاه بیزی پرداخته شد، که در پژوهش‌های بعدی می‌توان روش‌های تعیین اندازه‌ی نمونه را با در نظر گرفتن توزیع چوله نرمال برای اثرهای تصادفی توسعه بخشید.

جدول ۲- اندازه‌های نمونه‌ی مورد نیاز برای توان‌های بیزی مختلف در مطالعات طولی با پاسخ دودویی برای اندازه تکرارهای $T = ۳$

	$۱ - \beta$					
	۰/۸			۰/۹۵		
$۱ - \alpha$	۰/۹	۰/۹۵	۰/۹۹	۰/۹	۰/۹۵	۰/۹۹
T ۳	۲۱۰	۳۱۰	۳۲۰	۳۰۰	۳۹۰	۴۰۰
T ۵	۱۸۰	۲۱۰	۲۴۰	۱۹۰	۲۵۰	۳۰۰

توضیحات

1. Bayesian Power Criterion

مرجع‌ها

[1] Adcock, C.J. (1997). Sample Size Determination: a Review. Journal of the Royal Statistical Society: Series D (The Statistician), **46**, 261–283.

[2] Ahn, C., Heo, M. and Zhang, S. (2014). Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research. CRC Press.

[3] Baghfalaki, T. (2018). Bayesian Sample Size Determination for Longitudinal Studies with Continuous Response based on Different Scientific Questions of Interest. Journal of Biopharmaceutical Statistics, **29**, 1–27.

[4] Hedeker, D., Gibbons, R.D. and Waternaux, C. (1999). Sample Size Estimation for Longitudinal Designs with Attrition: Comparing Time-related Contrasts between Two Groups. Journal of Educational and Behavioral Statistics, **24**, 70–93.

[5] Kapur, K., Bhaumik, R., Tang, X.C. Hur, K., Reda. D.J. and Bhaumik, D.K. (2014). Sample Size Determination for Longitudinal Designs with Binary Response. Statistics in Medicine, **33**, 3781–3800.

[6] Lawrence, J., Wolfson, D.B. and Berger, R.D. (1995). Some Comments on Bayesian Sample Size Determination. Journal of the Royal Statistical Society: Series D (The Statistician), **44**, 167–171.

[7] Muller, K.E. and Barton, C.N. (1989). Approximate Power for Repeated-measures Anova Lacking Sphericity. Journal of the American Statistical Association, **84**, 549–555.

[8] Muller, K.E. and Lavange, L.M. (1992). Sharon Landesman Ramey, and Craig T Ramey. Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications. Journal of the American Statistical Association, **87**, 1209–1226.

[9] Muller, K.E. and Peterson, B.L. (1984). Practical Methods for Computing Power in Testing the Multivariate General Linear Hypothesis. Computational Statistics & Data Analysis, **2**, 143–158.

[10] Overall, J.E. and Doyle, S.R. (1994). Estimating Sample Sizes for Repeated Measurement Designs. Controlled Clinical Trials, **15**, 100–123.

[11] Pham-Gia, T. and Turkkan, N. (1992). Sample Size Determination in Bayesian Analysis. Journal of the Royal Statistical Society: Series D (The Statistician), **41**, 389–397.

[12] Wang, F. and Gelfand, A. E. (2002). A Simulation-based approach to Bayesian Sample Size Determination for Performance under a Given Model and for Separating Models. Statistical Science, **17**, 193–208.

مریم سوری
فوق لیسانس آمار
تهران، بزرگراه شهید چمران، خیابان یمن، میدان شهید شهریاری، بلوار دانشجو، دانشگاه شهید بهشتی، دانشکده‌ی علوم ریاضی.
رایانشانی: m.soori@mail.sbu.ac.ir

مجتبی گنجعلی
دکترای آمار
تهران، بزرگراه شهید چمران، خیابان یمن، میدان شهید شهریاری، بلوار دانشجو، دانشگاه شهید بهشتی، دانشکده‌ی علوم ریاضی.
رایانشانی: m-ganjali@sbu.ac.ir