

مجله علوم آماری، پاییز و زمستان ۱۳۹۵

جلد ۱۰، شماره ۲، ص ۲۲۱-۲۳۱

DOI: 10.18869/acadpub.jss.10.2.221

طرح D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دوم با اثرات تصادفی

حبیب جعفری، شیما پیرمحمدی

گروه آمار، دانشگاه رازی کرمانشاه

تاریخ دریافت: ۱۳۹۲/۸/۱۵ تاریخ آخرین بازنگری: ۱۳۹۵/۱/۲۲

چکیده: معیارهای بهینگی برای یافتن طرح بهینه مدل مورد بررسی، استفاده می‌شوند. این نوع مدل‌ها شامل مدل‌های مقایسه زوج شده هستند، که معیارهای بهینگی مقایسه‌های زوج شده بهینه را مشخص می‌کنند. در این مقاله علاوه بر معرفی مدل رگرسیونی درجه دوم با اثرات تصادفی، مدل‌های مقایسه زوج شده در این نوع مدل‌ها معرفی و طرح بهینه برای آن‌ها محاسبه شده است.

واژه‌های کلیدی: اثرات تصادفی، رگرسیون درجه دوم، طرح D -بهینه، مقایسه‌ی زوج شده.

۱ مقدمه

در بررسی بعضی مسائل استفاده از مقایسه زوج شده به جای بررسی مستقیم موضوع می‌تواند باعث دست یافتن به نتایج درست‌تری شود. به عنوان مثال برای بررسی این که از میان چند ماده شیرین کدام ماده شیرین‌تر است، در بررسی مستقیم، انتخاب شیرین‌ترین ماده علاوه بر این که دشوار است، ممکن است نتیجه درستی

آدرس الکترونیک مسئول مقاله: حبیب جعفری، h.jafari@razi.ac.ir

کد موضوع بندی ریاضی (۲۰۱۰): ۶۲K۰۵، ۶۲C۱۲

۲۲۲ طرح D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دو

را در پی نداشته باشد، در حالی که اگر مواد شیرین دو به دو مقایسه شوند، به راحتی می توان از میان دو ماده شیرین، ماده شیرین تر را انتخاب کرد (گراسهوف و همکاران، ۲۰۰۳).

مدل مقایسه زوج شده و به دست آوردن طرح بهینه در این مدل‌ها، توسط گروسمن و همکاران (۲۰۰۱) برای مدل‌های چند جمله‌ای انجام شده و کارایی طرح‌های بهینه نیز محاسبه شده است. استریت و همکاران (۲۰۰۱) در مدل طرح عاملی 2^k این موضوع را مورد بررسی قرار داده‌اند. از دیگر کارهای انجام گرفته در این زمینه می توان به گراسهوف و همکاران (۲۰۰۳) اشاره کرد که طرح بهینه برای مدل مقایسه زوج شده را در مدل طرح عاملی با اثر متقابل مرتبه اول به دست آورده و همین نویسندگان در سال ۲۰۰۴ این گونه طرح‌های بهینه را در مدل طرح عاملی با اثرات اصلی محاسبه کرده‌اند.

بختی که در این مقاله مورد نظر است، به کار بردن مقایسه زوج شده در مدل‌های آمیخته خطی و محاسبه طرح D -بهینه^۱ در چنین مدل‌هایی است. مدل آمیخته خطی مورد نظر در این مقاله، مدل رگرسیونی درجه دوم با اثرات تصادفی است که در بخش ۲ این مقاله علاوه بر معرفی مدل‌های آمیخته خطی، توضیحاتی در مورد معیار D -بهینگی و طرح‌های D -بهینه در این نوع مدل‌ها ارائه شده است. در بخش ۳ مدل مقایسه زوج شده معرفی می شود و طرح‌های D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دوم با اثرات تصادفی در بخش ۴ محاسبه شده و در نهایت بحث و نتیجه‌گیری در این رابطه در بخش آخر ذکر گردیده است.

۲ مدل‌های آمیخته خطی

مدل‌های آمیخته خطی به مدل‌های خطی گفته می شود که علاوه بر داشتن اثرات ثابت، شامل اثرات تصادفی نیز باشند. بنابراین مدل‌های آمیخته خطی دربرگیرنده مدل‌های خطی نیز هستند. در این حالت ساختار یک مدل آمیخته خطی را می توان

^۱ D-Optimal Design

به صورت

$$y_{ij} = \mathbf{f}^T(x_{ij})\boldsymbol{\gamma}_i + \varepsilon_{ij}; \quad i = 1, \dots, n, \quad j = 1, \dots, m_i \quad (1)$$

نشان داد، که در آن متغیر پاسخ مربوط به مشاهده j ام از i امین واحد^۲ جامعه و $\mathbf{f}(x_{ij})$ بردار تابع رگرسیونی است. اندیس i در این مدل نشان دهنده تعداد مشاهدات و اندیس j نشان دهنده تعداد تکرارهاست. در مدل (۱) داریم:

$$\boldsymbol{\gamma}_i = \boldsymbol{\beta} + \mathbf{b}_i; \quad \boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T, \quad \mathbf{b}_i = (b_{i1}, \dots, b_{ip})^T$$

که در آن بردار پارامترهای مدل (اثرات تثبیت شده^۳) و \mathbf{b}_i بردار اثرات تصادفی در مدل مورد نظر است. هم چنین

$$\mathbf{b}_i \sim N_p(0, \sigma_e^2 \mathbf{D}); \quad \mathbf{D} = \text{diag}\{d_1, \dots, d_p\}$$

که در آن $N_p(\cdot, \cdot)$ بیانگر توزیع نرمال p متغیره و $\text{diag}(d_1, \dots, d_p)$ نشان دهنده یک ماتریس قطری است که درایه های روی قطر اصلی آن به ترتیب d_1 و ... و d_p هستند و ε_{ij} ها نیز خطاهای تصادفی و مستقل از اثرات تصادفی هستند، به طوری که

$$\varepsilon_{ij} \sim N(0, \sigma_e^2).$$

۱.۲ معیار D -بهینگی و طرح بهینه در مدل های آمیخته خطی

طرح بهینه که برای دست یافتن به مناسب ترین برآوردگر به کار می رود، با توجه به معیارهای بهینگی که توابعی از ماتریس اطلاع فیشر هستند، محاسبه می شوند. (سیلوی، ۱۹۸۰). اگر $Y_i \sim f(y_i | \mathbf{x}_i, \boldsymbol{\beta})$ ، آنگاه ماتریس اطلاع فیشر به صورت

$$\mathbf{M}(\mathbf{x}_i) = -E\left(\frac{\partial^2 \ln f(y_i | \mathbf{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right). \quad (2)$$

^۲ Individual

^۳ Fixed Effects

۲۲۴ طرح D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دو

تعریف می‌گردد. حال اگر طرح دلخواه به فرم

$$\xi = \begin{pmatrix} x_1 & \dots & x_r \\ w_1 & \dots & w_r \end{pmatrix} \in \Xi$$

در نظر گرفته شود، که در آن

$$\Xi = \{ \xi \mid \sum_{i=1}^r w_i = 1, 0 \leq w_i \leq 1 \}$$

آن‌گاه ماتریس اطلاع مربوط به ξ به صورت محاسبه خواهد شد.

$$M(\xi) = \sum_{i=1}^r w_i M(x_i).$$

معیارهای بهینگی توابعی از ماتریس اطلاع فیشر $M(\xi)$ هستند. اما معیار بهینگی مورد استفاده در این مقاله معیار مشهور D -بهینگی است که به دلیل پایایی، بیشتر از سایر معیارهای بهینگی مورد استفاده قرار می‌گیرد (اتکینسون و همکاران، ۲۰۰۷). معیار D -بهینه براساس دترمینان ماتریس اطلاع به صورت

$$\psi(M(\xi)) = -\ln \det(M(\xi)); \quad \psi : M \rightarrow R$$

تعریف می‌شود که در آن $\det(\cdot)$ نشان‌دهنده دترمینان یک ماتریس M به صورت $M = \{M(\xi) \mid \xi \in \Xi\}$ تعریف می‌شود و R نیز مجموعه اعداد حقیقی است. بر همین اساس ξ^* یک طرح D -بهینه است، اگر

$$\xi^* = \arg \min_{\xi \in \Xi} (-\ln \det(M(\xi))).$$

اکنون با توجه به رابطه (۲) و مدل (۱) می‌توان ماتریس اطلاع فیشر را با فرض معلوم بودن D و σ_e^2 به صورت

$$M(x_{ij}) = \mathbf{f}(x_{ij}) V_{ij}^{-1} \mathbf{f}^T(x_{ij}) \quad (3)$$

محاسبه کرد، که در آن

$$V_{ij} = \sigma_e^2 (\mathbf{f}^T(x_{ij}) D \mathbf{f}(x_{ij}) + 1).$$

در بررسی طرح بهینه در مدل‌های آمیخته خطی، باید توجه کرد که هر واحد جامعه به طور مجزا مورد بررسی قرار می‌گیرد. اگر n واحد در جامعه وجود داشته باشد و ξ_i^* طرح بهینه مربوط به واحد i ام جامعه باشد، در این صورت خواهیم داشت:

$$\xi_i^* = \begin{pmatrix} x_{i1}^* & \dots & x_{ir_i}^* \\ w_{i1}^* & \dots & w_{ir_i}^* \end{pmatrix}; \quad i = 1, \dots, n. \quad (4)$$

هر یک از n طرح بهینه رابطه (۴) طرح بهینه مربوط به یکی از واحدهای جامعه است که در این حالت طرح بهینه مربوط به جامعه به صورت

$$\zeta^* = \begin{pmatrix} \xi_1^* & \dots & \xi_n^* \\ w_1^* & \dots & w_n^* \end{pmatrix}; \quad i = 1, \dots, n$$

خواهد بود. حال اگر $\xi_1^* = \dots = \xi_n^* = \xi^*$ ، یعنی طرح‌های بهینه برای هر واحد جامعه یکسان باشند، آن‌گاه طرح بهینه مربوط به جامعه عبارت خواهد بود از (اشمیلتر، ۲۰۰۷):

$$\zeta^* = \begin{pmatrix} \xi^* \\ 1 \end{pmatrix}$$

۳ مدل مقایسه زوج شده

با توجه به این که در این مقاله مقایسه زوج شده برای مدل رگرسیونی درجه دوم با اثرات تصادفی مد نظر است، بنابراین در این بخش مقایسه زوج شده در مدل (۱) معرفی می‌شود. لازم به ذکر است که در مدل‌های آمیخته خطی، مقایسه‌های زوج شده در هر واحد صورت می‌گیرند. برای این منظور با در نظر گرفتن مدل (۱)، مدل مقایسه زوج شده به صورت

$$y_i(j, j') = f_i^T(j, j')\gamma_i + \varepsilon_i(j, j'); \quad i = 1, \dots, n, j, j' = 1, \dots, m_i; j \neq j' \quad (5)$$

در نظر گرفته می‌شود. در این رابطه $y_i(j, j')$ متغیر پاسخ مربوط به مقایسه مشاهده‌های j ام و j' ام، در واحد i ام است و به صورت

$$y_i(j, j') = y_{ij} - y_{ij'}$$

۲۲۶ طرح D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دو

نمایش داده می شود. هم چنین $f_i(j, j')$ بردار تابع رگرسیونی است، که به صورت

$$f_i(j, j') = f(x_{ij}) - f(x_{ij'})$$

محاسبه می شود و $\varepsilon_i(j, j')$ نیز خطای تصادفی مدل است که

$$\varepsilon_i(j, j') = \varepsilon_{ij} - \varepsilon_{ij'}$$

در مدل (۵) صرف نظر از توزیع ε_{ij} ها، فرض شده است که

$$\varepsilon_i(j, j') \sim N(0, \sigma^2)$$

و هم چنین $\varepsilon_i(j, j')$ ها علاوه بر این که از هم مستقل اند، از اثرات تصادفی موجود در مدل نیز مستقل می باشند. اکنون با توجه به مدل (۵) و رابطه (۳) ماتریس اطلاع فیشر به صورت

$$M(x_{ij}, x_{ij'}) = f_i(j, j') V_i^{-1}(j, j') f_i^T(j, j') \quad (6)$$

محاسبه می شود، که در آن

$$V_i(j, j') = \sigma^2 (f_i^T(j, j') D f_i(j, j') + 1).$$

در مدل مقایسه زوج شده ماتریس اطلاع مربوط به زوج (j, j') برابر ماتریس اطلاع مربوط به زوج (j', j) است (گراسهوف و همکاران، ۲۰۰۴)، یعنی

$$M(j, j') = M(j', j)$$

۴ طرح D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دوم با اثرات تصادفی

در این بخش طرح D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دوم با شیب تصادفی و طرح D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دوم

با انحنای تصادفی محاسبه شده است. برای دست یافتن به مدل مقایسه زوج شده در رگرسیون درجه دوم با اثرات تصادفی، لازم است مدل رگرسیونی درجه دوم با اثرات تصادفی ارائه شود. مدل رگرسیونی درجه دومی که در این مقاله معرفی می شود، بدون عرض از مبدأ در نظر گرفته شده است. در مدل مقایسه زوج شده عرض از مبدأ از بین رفته و باعث صفر شدن دترمینان ماتریس اطلاع می شود. بر همین اساس مدل رگرسیونی درجه دوم بدون عرض از مبدأ در نظر گرفته شده است، به طوری که در مدل های (۱) و (۵) خواهیم داشت:

$$f(x_{ij}) = (x_{ij}, x_{ij}^2)^T, \quad \beta = (\beta_1, \beta_2)^T, \quad b_i = (b_{i1}, b_{i2})^T$$

که

$$x_{ij} \in [-5, 5]; \quad j = 1, \dots, m_i.$$

اکنون مدل مقایسه زوج شده در رابطه (۵) را با مفروضات بالا در نظر گرفته و برای حالت های شیب تصادفی و انحنای تصادفی، به محاسبه طرح D -بهینه می پردازیم.

۱.۴ رگرسیون درجه دو با شیب تصادفی

مدل رگرسیونی درجه دو با شیب تصادفی به صورت مدل (۱) و با ویژگی هایی به صورت

$$f(x_{ij}) = (x_{ij}, x_{ij}^2)^T, \quad \beta = (\beta_1, \beta_2)^T, \quad b_i = (b_{i1}, 0)^T$$

است و هم چنین

$$x_{ij} \in [-5, 5]; \quad j = 1, \dots, m_i$$

بنابراین با توجه به بخش ۳ مدل مقایسه زوج شده در رگرسیون درجه دو با شیب تصادفی به صورت

$$y_i(j, j') = f_i^T(j, j')\gamma_i + \varepsilon_i(j, j'); \quad i = 1, \dots, n, \quad j, j' = 1, \dots, m_i; \quad j \neq j'$$

۲۲۸ طرح D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دو

خواهد بود، که در آن

$$\varepsilon_i(j, j') \sim N(0, \sigma^2), \quad b_i \sim N_2(0, \sigma^2 D); \quad D = \text{diag}\{d_1, 0\}$$

حال با فرض این که $\sigma^2 = 1$ و D معلوم باشد، با توجه به رابطه (۶) ماتریس اطلاع فیشر به صورت

$$M(x_{ij}, x_{ij'}) = \frac{(x_{ij} - x_{ij'})^2}{d_1(x_{ij} - x_{ij'})^2 + 1} \begin{pmatrix} 1 & x_{ij} + x_{ij'} \\ x_{ij} + x_{ij'} & (x_{ij} + x_{ij'})^2 \end{pmatrix}$$

به دست می آید. اکنون اگر به طور خاص برای واحد نام جامعه طرح به صورت

$$\xi_i = \begin{pmatrix} 1 \\ x \\ \frac{w}{1-w} \end{pmatrix}$$

در نظر گرفته شود، آن گاه دترمینان ماتریس اطلاع فیشر مربوط به آن به صورت

$$\det(M(\xi_i)) = \frac{w^2(1-x^2)^2(1-x)^2}{(d_1(1-x)^2 + 1)^2} + \frac{4w(1-w)(1-x^2)^2}{(4d_1 + 1)(d_1(1-x)^2 + 1)}$$

خواهد بود، که با ماکسیمم کردن آن نسبت به x و w و به ازای مقادیر مختلف d_1 ، طرح های D -بهینه مورد نظر معین می شوند. مقادیر بهینه x و w که به ترتیب با x^* و w^* نشان داده می شوند، به ازای مقادیر مختلف d_1 در جدول ۱ ارائه شده است.

جدول ۱: مقادیر بهینه x و w به ازای مقادیر مختلف d_1

d_1	x^*	w^*
0	-0/۲۳۶۰	0/۸۰۹۰
0/0۳	-0/۲۳۶۰	0/۸۴۶۰
0/0۶	-0/۲۳۶۰	0/۸۸۳۰
0/0۹	-0/۲۳۶۰	0/۹۲۰۲
0/۱۲	-0/۲۳۶۰	0/۹۵۷۲
0/۱۵	-0/۲۳۶۰	0/۹۹۴۴

همان طور که ملاحظه می شود با افزایش واریانس اثر تصادفی، تاثیر نقاط $(-0/۲۳۶۰)$ و $(0/۲۳۶۰)$ نیز در حال افزایش است و نقاط طرح با تغییر d_1 ، تغییر نمی کنند. به

ازای $d_1 = 0$ طرح D -بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دو با اثرات ثابت به دست آمده است (گروسمن و همکاران، ۲۰۰۱). حال چون واریانس اثر تصادفی برای هر واحد جامعه یکسان است، بنابراین $\xi_1^* = \dots = \xi_n^*$ پس طرح D -بهینه مربوط به جامعه در این حالت به صورت $\xi^* = \begin{pmatrix} \xi_i^* \\ 1 \end{pmatrix}$ خواهد بود.

۲.۴ رگرسیون درجه دو با انحنا تصادفی

برای مدل مقایسه زوج شده در رگرسیون درجه دو با انحنا تصادفی، $b_i = (0, b_{i2})^T$ است. یعنی این مدل مانند مدل (۵) خواهد بود، با این تفاوت که در این حالت $D = \text{diag}\{0, d_2\}$ معلوم و $\sigma^2 = 1$ است، در این صورت ماتریس اطلاع فیشر به صورت

$$M(x_{ij}, x_{ij'}) = \frac{(x_{ij} - x_{ij'})^2}{d_2(x_{ij}^2 - x_{ij'}^2)^2 + 1} \begin{pmatrix} 1 & x_{ij} + x_{ij'} \\ x_{ij} + x_{ij'} & (x_{ij} + x_{ij'})^2 \end{pmatrix}.$$

محاسبه می شود، اکنون اگر به طور خاص برای واحد نام جامعه طرح

$$\xi_i = \begin{pmatrix} \begin{pmatrix} 1 \\ x \\ w \end{pmatrix} & \begin{pmatrix} -1 \\ -x \\ w \end{pmatrix} & \begin{pmatrix} 1 \\ -1 \\ 1-w \end{pmatrix} \end{pmatrix}$$

در نظر گرفته شود، آن گاه

$$\det(M(\xi_i)) = \frac{w^2(1-x^2)^2(1-x)^2}{(d_2(1-x^2)^2 + 1)^2} + \frac{4w(1-w)(1-x^2)^2}{d_2(1-x^2)^2 + 1}$$

که با ماکسیمم کردن دترمینان نسبت به x و w و به ازای مقادیر مختلف d_2 طرح های D -بهینه معین می شوند. مقادیر بهینه x و w که به ترتیب با x^* و w^* نشان داده می شوند، در جدول ۲ ارائه شده اند. همان طور که ملاحظه می شود از مقدار $d_2 = 0/02$ به بعد اگرچه نقاط طرح ثابت هستند، اما با افزایش d_2 ، وزن نقاط $(4/2360)$ و $(-4/2360)$ در حال کاهش و وزن نقطه ی (-1) در حال افزایش است. طرح D -بهینه در جدول ۲ به ازای $d_2 = 0$ و در جدول ۱، به ازای $d_1 = 0$ به دست آمده است.

۲۳۰ طرح D - بهینه برای مدل مقایسه زوج شده در رگرسیون درجه دو

جدول ۲: مقادیر بهینه x و w به ازای مقادیر مختلف d_2

d_2	x^*	w^*
۰	-۰/۲۳۶۰	۰/۸۰۹۰
۰/۰۲	۴/۲۳۶۰	۰/۸۱۷۴
۰/۱	۴/۲۳۶۰	۰/۵۴۸۲
۱	۴/۲۳۶۰	۰/۵۰۴۴
۲	۴/۲۳۶۰	۰/۵۰۲۲
۱۰	۴/۲۳۶۰	۰/۵۰۰۴
۱۰۰	۴/۲۳۶۰	۰/۵۰۰۰

اکنون با توجه به ثابت بودن واریانس اثر تصادفی در هر واحد جامعه داریم
 $\xi_1^* = \dots = \xi_n^*$. بنابراین طرح D - بهینه مربوط به جامعه به صورت $\zeta^* = \begin{pmatrix} \xi_i^* \\ 1 \end{pmatrix}$ خواهد بود.

بحث و نتیجه گیری

طرح D - بهینه که برای مدل مقایسه زوج شده در مدل رگرسیونی درجه دو به کار رفته، در دو حالت محاسبه شده است: شیب تصادفی و انحنا تصادفی. در هر دو حالت، مدل های رگرسیونی درجه دو بدون عرض از مبدأ در نظر گرفته شده است. در حالت اول با افزایش واریانس اثر تصادفی، نقاط طرح یکسانند، اما تأثیر نقاطی که دارای وزن یکسان هستند نیز افزایش یافته و به همین نسبت، تأثیر نقطه ای که دارای وزن متفاوت است، کاهش می یابد. در حالت دوم با افزایش واریانس اثر تصادفی تأثیر نقاطی که دارای وزن یکسانند، کاهش یافته و تأثیر نقطه ای که دارای وزن متفاوت بوده، در حال افزایش است. بنابراین وقتی شیب تصادفی باشد، تأثیر نقاط هم وزن با افزایش واریانس اثر تصادفی، افزایش خواهد یافت. این در حالی است که اگر انحنا، تصادفی باشد، تأثیر نقاط هم وزن با افزایش واریانس اثر تصادفی کاهش می یابد. به عنوان یک پیشنهاد می توان در تحقیقات آتی، شیب و انحنا را به طور هم زمان، تصادفی در نظر گرفت و نتایج مربوط به آن را بررسی و با حالت هایی که فقط شیب یا فقط انحنا، تصادفی باشد، مقایسه کرد.

- Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007), *Optimum Experimental Designs, with SAS*, Oxford University Press, New York.
- Grasshoff, U., Grossmann, H., Holling, H. and Schwabe, R. (2004), Optimal Designs for Main Effects in Linear Paired Comparison Models, *Journal of Statistical Planning and Inference*, **126**, 361-376.
- Grasshoff, U., Grossmann, H., Holling, H. and Schwabe, R. (2003), Optimal Paired Comparison Designs for First-Order Interactions, *Statistics*, **37**, 373-386.
- Grossmann, H., Holling, H., Grasshoff, U. and Schwabe, R. (2001), *Efficient Designs for Paired Comparisons with a Polynomial Factor*, In Atkinson, A., Bogacka, B., Zhigljavsky. A. (Eds), *Optimum Design 2000*, Kluwer Academic Publishers, Dordrecht, 45-56.
- Schmelter, T. (2007), The Optimality of Single-Group Designs for Certain Mixed Models, *Metrika*, **65**, 183-193.
- Silvey, S. D. (1980), *Optimal Design an Introduction to the Theory for Parameter Estimation*, Chapman and Hall, London.
- Street, D. J., Bunch, D. S. and Moore, B. J. (2001), Optimal Designs for 2^k Paired Comparison Experiments, *Communications in Statistics-Theory and Methods*, **30**, 2149-2171.