بسم الله الرحمن الرحيم

کلیه حقوق مادی مترتب بر نتایج مطالعات، ابتکارات و
نوآوری‌های ناشی از تحقیق موضوع این پایان‌نامه
متعلق به دانشگاه رازی است.

دانشکده علوم پایه
گروه آمار

پایان‌نامه جهت اخذ درجه کارشناسی ارشد
رشته آمار ریاضی

عنوان :

طرح D - بهینه موضعی برای مدل رگرسیون لجستیک با سه متغیر مستقل

استاد راهنما:
دکتر حبیب جعفری

نگارش:
مرضیه ظاهری

اسفند ۱۳۹۱

دانشکده علوم پایه
گروه آمار

پایان‌نامه جهت اخذ درجه کارشناسی ارشد
رشته آمار ریاضی

نام دانشجو:
مرضیه ظاهری

تحت عنوان :

طرح D - بهینه موضعی برای مدل رگرسیون لجستیک با سه متغیر مستقل

در تاریخ توسط هیأت داوران زیر بررسی و با درجه به تصویب نهایی رسید.

استاد راهنمای پایان‌نامه دکتر حبیب جعفری با مرتبه‌ی علمی امضاء:

استاد داور دکتر مهرداد نیاپرست با مرتبه‌ی علمی امضاء:

استاد داور دکتر سلیمان خزائی با مرتبه‌ی علمی امضاء:

سپاس و قدردانی

ای بی‌نهایت آشنا، ای خلقت آفرین، از آن روز که مرا آفریدی ارزش آن را پیدا کردم که معنای بودن را تجربه کنم و این اساس بندگی من کوچک در برابر دانش خدایی توی بزرگ بود. سپاس تو را که روح انسان را از تلاش و حکمت آفریدی و بدین راه، راهنمایی‌اش کردی. هر آن‌چه دارم از تو و هر آن‌چه بدست خواهم آورد نیز از تو.

سپاسگزار کسانی هستم که سرآغاز تولد منند. از یکی زاده می‌شوم و از دیگری جاودانه. پدر و مادری که تار مویی از آن‌ها به پای من، سیاه نماند، و اساتیدی که سپیدی را برتخته سیاه زندگیم نگاشتند.

سپاسگزار استادی هستم که اندیشیدن را به من آموخت، نه اندیشه‌ها را، به من آموخت که برای رسیدن باید فکر کرد، ایستاد بر فراز قله‌های معرفت و افق‌های باز را لمس کرد. جناب دکتر حبیب جعفری.

بر خود واجب می‌دانم از زحمات اساتید گرانقدرم، آقایان دکتر مهرداد نیاپرست و دکتر سلیمان خزائی که به‌عنوان داوران در این تحقیق با مطالعه و ارایه‌ی نظرات سودمند به داوری نشستند، قدردانی کنم.

سپاس بیکران برهمدلی، همراهی و همگامی همسر عزیزم داود، خواهر مهربانم راضیه، برادر دلسوزم رضا، پدر و مادر همسرم که وجودشان آسمان زندگی‌ام را فروغی صد چندان بخشید و همواره چون کوهی استوار پشتیبان و حامی من بودند و بسیاری از سختی‌ها را برایم آسان‌تر نمودند و بدون یاری آن‌ها اتمام این پایان‌نامه امکان‌پذیر نبود.

مرضیه ظاهری
اسفند ۱۳۹۱

تقدیم به چهار وجود مقدس:

روح مادرم به زلالی چشمه

پدرم به استواری کوه

همسرم به صمیمیت باران

و استاد گرانقدرم جناب دکتر حبیب جعفری

چکیده

به دلیل کاربرد فراوان مدل‌های غیرخطی در زمینه‌های مختلف، این مدل‌ها از اهمیت خاصی برخوردار می‌باشند. در این پایان‌نامه طرح‌های بهینه برای مدل‌های خاصی از این نوع (مدل رگرسیون لجستیک) مورد بررسی قرار گرفته است. بر همین اساس با توجه به اینکه معیار D- بهینگی تابعی از ماتریس اطلاع است و در مدل‌های غیرخطی ماتریس اطلاع وابسته به پارامترهای مدل می‌باشد بنابراین از معیار D- بهینگی موضعی برای بدست آوردن طرح بهینه موضعی استفاده شده است. در این پایان‌نامه با توجه به معیار بهینه ذکر شده، طرح مناسب برای مدل رگرسیون لجستیک با سه متغیر مستقل معرفی و بر اساس فضای پارامتر، طرح‌های D- بهینه موضعی محاسبه شده‌اند.

کلمات کلیدی: *معیار D- بهینه، طرح D- بهینه موضعی، ماتریس اطلاع، قضیه هم‌ارزی، رگرسیون لجستیک*

فهرست مطالب

عنوان صفحه

۱ تعاریف و اصطلاحات ۱

۱-۱ مقدمه ۲

۱-۲ طرح ۲

۱-۲-۱ طرح آزمایش ۲

۱-۳ طرح بهینه ۴

۱-۴ مدل ۵

۱-۴-۱ مدل‌های رگرسیون خطی ۵

۱-۴-۲ مدل های رگرسیون غیرخطی ۵

۱-۴-۳ مدل‌های خطی تعمیم‌یافته ۶

۱-۴-۴ ماتریس اطلاع فیشر ۶

۱-۴-۵ ماتریس اطلاع طرح مربوط به مدل‌های خطی و خصوصیات آن ۷

۱-۵ معیارهای بهینه سازی ۸

۱-۵-۱ ویژگی‌های معیارهای بهینه‌سازی ۱۰

۱-۵-۲ مشتق‌های جهتی ۱۱

۱-۵-۳ ویژگی‌های مشتق جهتی ۱۲

۱-۵-۴ قضیه هم‌ارزی ۱۳

۲ طرح D - بهینه برای مدل رگرسیون خطی ۱۸

۲-۱ مقدمه ۱۹

۲-۲ معرفی مدل ۱۹

۲-۲-۱ مدل رگرسیون خطی ۱۹

۲-۲-۲ برآورد پارامترها به روش کمترین مربعات ۲۰

۲-۲-۳ ماتریس اطلاع طرح ۲۱

۲-۳ طرح D - بهینه برای مدل رگرسیون خطی ساده ۲۲

۲-۴ طرح D - بهینه برای مدل رگرسیون خطی چندگانه با دو رگرسور ۲۵

۲-۵ طرح D-بهینه برای مدل رگرسیون چندجمله‌ای ۲۷

۳ طرح‌های D-بهینه موضعی برای مدل رگرسیون لجستیک بادو متغیرمستقل . . ۲۹

۳-۱ مقدمه ۳۰

۳-۲ رگرسیون لجستیک ۳۱

۳-۲-۱ برآورد درستنمایی پارامترها ۳۱

۳-۳ مدل رگرسیون لجستیک با دو متغیر مستقل ۳۳

۳-۴ طرح‌های D-بهینه ۳۶

۳-۴-۱ طرحی مبنی بر چهار نقطه ۳۶

۳-۴-۲ طرحی مبنی بر سه نقطه ۳۹

۳-۴-۳ چندین طرح دیگر مبنی بر چهار نقطه ۴۱

۳-۴-۴ بررسی کارایی طرح‌های چهار نقطه‌ای ۴۴

۴ طرح D-بهینه موضعی برای مدل رگرسیون لجستیک با سه متغیر مستقل . . . ۴۷

۴-۱ مقدمه ۴۸

۴-۲ مدل رگرسیون لجستیک با سه متغیر مستقل ۴۸

۴-۳ نتیجه‌گیری ۵۲

منابع و مآخذ ۵۴

فهرست نشانه‌ها و نمادها

ξ	طرح آزمایش
ξ^*	طرح آزمایش بهینه
Ξ	فضای طرح آزمایش
M	ماتریس اطلاع
arg	آرگومان
D	معیار D- بهینگی
D_{eff}	معیار D- کارایی
A	معیار A- بهینگی
E	معیار E- بهینگی
G	معیار G- بهینگی
ψ	تابع معیار بهینگی
det	دترمینان
d	تابع حساسیت
α	مقدار ثابت
tr	مجموع عناصر روی قطر اصلی
β	پارامتر مدل
var	واریانس
∂	مشتق جزئی
ℓ	لگاریتم درستنمایی
$L_{p(x)}$	p امین چندجمله‌ای لژاندر x

پیشگفتار

طرح آزمایش‌ها یکی از مباحث مهم کاربردی علوم آماری است که در علوم و صنایع مختلف نقش مهمی را ایفا می‌کند. در این پایان‌نامه شاخه‌ای از این مبحث، به نام طرح آزمایش‌های بهینه بررسی خواهد شد. انگیزه‌ی پایه‌ای برای طرح آزمایش‌های بهینه، پیدا کردن طرحی ایده‌ال است که تحت اجرای آزمایش با این طرح، استنباط‌های مناسبی در مورد پارامترهای مدل به دست آید.

در مباحث آماری، در حالت کلی مدل‌ها به دو دسته، مدل‌های خطی و غیرخطی تقسیم می‌شوند؛ مدل‌های غیرخطی به‌طور فراوان، در زمینه‌هایی همچون پزشکی و داروسازی، مورد استفاده قرار می‌گیرند. هدف اصلی این پایان‌نامه بررسی وضعیت طرح‌های بهینه برای این نوع مدل‌ها است. اگر چه طرح‌های بهینه برای مدل‌های خطی از سال‌ها پیش مورد بررسی قرار گرفته و نتایج زیادی در این مورد بدست آمده است، ولی برای مدل‌های غیرخطی به دلیل پیچیدگی کار با این مدل‌ها، نتایج و دستاوردهای کمی حاصل شده است، به همین خاطر ما را برآن داشت که مطالعاتی را در رابطه با این نوع مدل‌ها انجام داده، طرح‌های مناسب را تعریف کرده و بهینه آن‌ها را بدست آوریم، بنابراین این پایان‌نامه به صورت زیر دسته بندی شده است:

در فصل اول این پایان‌نامه، طرح بهینه و معیارهای بهینه‌سازی تعریف می‌شوند. هم‌چنین طرح‌های بهینه برای مدل‌های رگرسیون خطی در فصل دوم مورد بررسی قرار گرفته‌اند. فصل سوم شامل بحث در رابطه با طرح‌های بهینه برای مدل رگرسیون لجستیک با دو متغیر مستقل می‌باشد، چون ماتریس اطلاع برای این نوع مدل‌ها وابسته به مقادیر نامعلوم پارامترها است بنابراین برخلاف بحث بدست آوردن طرح‌های بهینه در مدل‌های خطی، محاسبه طرح‌های بهینه در این نوع مدل‌ها از پیچیدگی خاصی برخوردار می‌باشد. رویکردهای مختلفی برای غلبه بر این مشکل در ادامه‌ی فصل، مطرح شده است.

لازم به ذکر است که دو فصل دوم و سوم این پایان‌نامه، مربوط به معرفی طرح‌های بهینه است و در نهایت در فصل چهارم طرح مناسب برای مدل رگرسیون لجستیک با سه متغیر مستقل (در حالت خاص) معرفی و با توجه به معیار D- بهینه، طرح D- بهینه موضعی برای آن محاسبه گردیده است. در فصل پنجم نتیجه‌گیری از چهار فصل اول ارائه شده است.

فصل اول

تعاریف و اصطلاحات

۱-۱ مقدمه

در این فصل مفاهیم اولیه مربوط به طرح‌های بهینه، معیارهای بهینه‌سازی و قضایای مربوطه شرح داده شده‌اند.من

۱-۲ طرح

بحث کیفیت در دهه‌های اخیر شامل مباحثی چون کنترل آماری فرآیند، مدیریت کیفیت جامع، مهندسی مجدد، بهبود مستر ژاپنی (کایزن)، تحلیل هزینه‌های کیفیت، بهبود مستمر فرآیند است.

تکنیک طراحی آزمایش‌ها [۱] نیز یکی از تکنیک‌های بهبود کیفیت است که در دهه‌های ۱۹۸۰ و ۱۹۹۰ به عنوان یک مزیت رقابتی در کشورهای غربی و ژاپن مطرح شده است. استفاده صحیح از روش‌های طراحی آزمایش‌های آماری می‌تواند باعث سهولت در مراحل طراحی و تولید محصولات جدید و بهبود محصولات موجود گردد. این اصول در اغلب صنایع نظیر صنایع الکترونیک و نیمه هادی، هوافضا، اتومبیل، تجهیزات پزشکی، غدایی، داروسازی و صنایع شیمیایی و فرآیندی به کار گرفته شده‌اند. باید به این نکته توجه داشت که ابزارهای کیفیت فقط دارای توانایی اعلام عیب بوده و اینکه چه عواملی و چگونه بر روی کیفیت فرآیند اثر می گذارند، اگر این عوامل تغییر کنند کیفیت چگونه تغییر می‌کند؟ این عوامل چگونه تغییر داده شوند تا بهترین عملکرد کیفی بدست آید؟ عوامل کنترل‌پذیر چگونه باشند تا اثر عوامل غیرقابل کنترل بر روی کیفیت حداقل شود؟ این‌ها سؤالاتی هستند که با انجام آزمایش می‌توان بدانها پاسخ داد.

۱-۲-۱ طرح آزمایش

در کلیه زمینه‌های تحقیق، محققین از آزمایش‌ها با هدف کسب اطلاعاتی در مورد یک فرآیند یا یک سیستم خاص استفاده می‌کنند. اصطلاحاً یک آزمایش [۲] یک آزمون [۳] است. به طور دقیق‌تر، یک آزمایش را می‌توان یک آزمون یا مجموعه‌ای از آزمون‌هایی تعریف کرد که به طور هدفمند تغییراتی در متغیرهای ورودی یک فرآیند یا یک سیستم ایجاد می‌کنند تا از این طریق، امکان مشاهده و شناسایی

---

[۱] Design of Experiments

[۲] Experiment

[۳] Test

تغییرات ممکن در پاسخ خروجی فراهم شود. انجام آزمایش همواره متضمن هزینـه و زمان است. از این رو انجام آزمایش‌های مؤثـر که با صرف حداقل هزینه و زمان، بیشترین اطلاعات را بدست بدهند آرمان هر محقق است. اما با افزایش تعداد عوامل، هزینه و زمان نیز به صورت صعودی افزایش پیدا می‌کنند. بنابراین به روشی نیاز است که در آن بتوان با صرف حداقل هزینه و زمان، به بیشترین اطلاعات در مورد فرآیند دست پیدا کرد، نتیجه‌گیری‌های منطقی ارائه داد و مدارك مستند در خصوص فرآیند بدست آورد. روشی که به بهترین شکل اهداف ذکر شده را برآورده می سازد، طرح آزمایش‌ها نام دارد. به عبارتی طرح آزمایش، عبارتست از الگوی عمل برای انجام یك آزمایش (هر عملی که منجر به جمع‌آوری اطلاعات گردد) به طوریکه بتوان با استفاده از مشاهدات و روشهای آماری نتیجه‌گیری معتبر به عمل آورد. با توجه به اینکه عوامل قابل کنترل بوسیله انتخاب طرح مناسب کنترل می‌شوند، عوامل غیرقابل کنترل نیز با انتخاب مدل مناسب و فرضیاتی روی مشاهدات آزمایشی(میزان محصول) سعی در به حداقل رساندن تغییرات آنها می‌نماید.

در هر آزمایش تعدادی از متغیرها مقادیر آن‌ها تحت کنترل تحلیل‌گر بوده و آنها را با توجه به نیاز آزمایشی که انجام می‌دهـد انتخاب می‌کند. به همین دلیل به این متغیرها، متغیرهای تحت کنترل نیز گفته می‌شود. متغیرهای غیرقابل کنترل موجب ایجاد پارازیت[1] در فرآیند شده و امکان شناسایی متغیرهای قابل کنترل را کاهش می‌دهند. لذا شناسایی این متغیرها نیز از اهمیت خاصی برخوردار بوده و باید نسبت به جمع‌آوری اطلاعات در مورد آنها اقدام شود.

هدف از طرح آزمایش‌ها:

۱) کسب حداکثر اطلاعات و شناخت از فرآیندها

۲) صرف حداقل هزینه‌ها

۳) کسب حداقل زمان برای اجرای فرآیندها

دستاوردهای طرح آزمایش‌ها:

۱) شناخت مشخصه‌های محصول و فرآیند و تعیین موثرترین و مهم‌ترین عوامل کنترل فرآیند.

۲) تعیین نحوه تاثیر ورودی‌های فرآیند (یا مشخصه‌های کنترلی) بر روی خروجی فرآیند و محصول (یا مشخصه‌های کیفی یا عملکردی) در کوتاهترین زمان و با کمترین هزینه ممکن.

۳) تعیین متغیرهای ورودیی که بیشترین تأثیر را در خروجی محصول یا فرآیند دارند.

---

[1] Noise

۴) مدل‌سازی فرآیند و تعیین رابطه ورودی‌ها و خروجی‌های (محصول یا عملکرد) فرآیند و اقدام جهت بهبود مستمر محصول و فرآیند، با کمترین هزینه و در کوتاهترین زمان.

۵) بهینه‌سازی و افزایش راندمان و بازدهی فرآیندها.

۶) تعیین مقدار دقیق پارامترهای ورودی به طوری که تقریباً خروجی فرآیند همیشه به هدف نزدیك باشد.

۷) تعیین مقدار دقیق پارامترهای ورودی به طوری که تغییرپذیری خروجی فرآیند کوچك باشد.

۸) کاهش تغییرپذیری در فرآیندی که شامل تعیین مقدار دقیق پارامترهای ورودی قابل کنترل است، به نحوی که اثر عوامل کنترل‌ناپذیر مینیمم شود.

۹) تدوین دانش فنی و شناخت عملکرد محصول و فرآیند با توجه به پارامترهای ورودی آنها و اقدام جهت کاهش موثر هزینه و زمان توسعه و بهبود محصول و فرآیند.

۱-۳ طرح بهینه

طرح‌های بهینه کلاسی از طرح آزمایش‌ها هستند که نسبت به برخی از معیارها بهینه‌اند.

مزایای طرح بهینه:

۱) چنین طرح هایی باعث می‌شوند، پارامترها بدون اریبی و با کمترین واریانس برآورد شوند.

۲) طرح‌های بهینه، با استفاده از مدل‌های آماری که با اجرای تعداد کمتری از آزمایش‌ها، برآورد می‌شوند، منجر به کاهش هزینه‌های آزمایش می‌شوند.

۳) حتی زمانی که فضای طرح محدود شده، می‌توان طرح بهینه را بدست آورد.

حال با توجه به موارد ذکر شده‌ی فوق طرح ξ با m نقطه پشتیبانی[۱] به صورت

$$\xi = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ w_1 & w_2 & \dots & w_m \end{pmatrix} \in \Xi \qquad (۱-۱)$$

تعریف می‌شود که شامل ۲ نوع مؤلفه است؛ مؤلفه اول، x_i ها، معرف نقاط پشتیبان بوده که متعلق به فضای طرح بوده و شامل متغیرهای تشریح‌کننده‌اند و شرایط آزمایشی را توصیف می‌کنند. مؤلفه دوم، w_i ها، نشان‌دهنده وزن‌ها که متناسب با مؤلفه اول بوده که هرکدام بزرگتر مساوی صفر بوده و مجموع

[۱] support point

آن‌ها برابر ۱ است.
با توجه به اینکه ξ یک اندازه احتمال است بنابراین
$$\int_X \xi(dx) = ۱.$$

۱-۴ مدل

۱-۴-۱ مدل‌های رگرسیون خطی

در این پایان‌نامه، مدل‌های رگرسیون خطی به مدل‌هایی گفته می‌شود که ارتباط بین متغیرهای مستقل و پارامترهای مدل به صورت خطی است و به شکل زیر نمایش داده می‌شوند:

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

که $\mathbf{f}(x_i) = (۱ \quad x_{۱i}, \cdots, x_{pi})^T$ ماتریس $\mathbf{F} = (\mathbf{f}^T(x_۱), \cdots, \mathbf{f}^T(x_n))^T$ ای از مشاهدات است و $n \times p$ و $\boldsymbol{\beta} = (\beta_۱, \cdots, \beta_p)^T$، بردار $p \times ۱$ ای از پارامترهاست، عناصر $\boldsymbol{\varepsilon} = (\epsilon_۱, \cdots, \epsilon_n)^T$، جملات باقیمانده خطا هستند که معمولاً فرض می‌شود به صورت نرمال با میانگین صفر و واریانس ثابت $\sigma^۲$ توزیع شده‌اند که از هم مستقل‌اند، یکی از رایج‌ترین مثال‌ها مدل خطی چندجمله‌ای درجه دو می‌باشد.

۱-۴-۲ مدل های رگرسیون غیرخطی

یکی از عوامل علاقه‌مندی در بدست‌آوردن طرح‌های بهینه، مدل‌های غیرخطی‌اند و به صورت زیر تعریف می‌شوند

$$\mathbf{y} = \eta(\mathbf{x}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

که $\mathbf{x}$ ماتریس $n \times p$، $\boldsymbol{\beta}$ هم بردار پارامترهای $p \times ۱$ است و عناصر $\boldsymbol{\varepsilon}$ هم نوعاً فرض شده است که دارای توزیع نرمال با میانگین صفر و واریانس $\sigma^۲$ هستند و ناهمبسته‌اند. این چنین مدل‌هایی تعمیمی از مدل رگرسیون خطی بوده که دیگر محدود به این فرض نیستند که در پارامترها خطی باشند، مثالی خاص از مدل‌های رگرسیون غیرخطی را می‌توان، مدل میکائیل- منتن [۱] سه پارامتری (مربع)، با توجه به یک متغیر مستقل می‌باشد که شکل این مدل عبارت است از:

$$\eta(\mathbf{x}, \boldsymbol{\beta}) = \frac{\beta_۱ x}{\beta_۲ + x + \beta_۳ x^۲} + \epsilon$$

[۱] Michaelis-Menten

۱-۴-۳ مدل‌های خطی تعمیم‌یافته

مدل‌های خطی تعمیم‌یافته [۱] شامل کلاس بزرگی از مدل‌هایی هم‌چون مدل‌های رگرسیون خطی ساده، مدل‌های پاسخ دودویی و مدل داده بقا می‌باشند. آن‌ها را می‌توان به عنوان کلاس واحدی درنظر گرفت. ساختن یک مدل خطی تعمیم‌یافته شامل سه تصمیم زیر است:

۱) داده‌ها از چه توزیعی تبعیت می‌کنند.

۲) چه تابعی از میانگین به عنوان پیش‌بینی کننده خطی مدل است.

۳) چه چیزی پیش‌بینی می‌شود.

الف) توزیع y

به طور معمول فرض شده است که بردار y شامل اندازه‌های مستقلی از توزیعی با چگالی از خانواده نمایی یا شبه‌نمایی است یعنی

$$y_i \sim indep. f_{Y_i}(y_i) f_{Y_i}(y_i) = e^{[y_i\gamma_i - b(\gamma_i)]/\tau^2 - c(y_i,\tau)} \qquad (۱-۲)$$

که برای راحتی توزیع به شکل کانونی نوشته می‌شود.

ب) تابع پیوند

به طور معمول هدف برقرار کردن ارتباط بین پارامترهای توزیع با پیش‌بینی کننده‌های مختلف است که این کار از طریق مدل‌بندی تبدیلی از میانگین یعنی π_i انجام می‌شود:

$$E(y_i) = \pi_i g(\pi_i) = x_i^T \beta$$

که $g(.)$، تابعی معلوم استو تابع پیوند نامیده می‌شود (زیرا پیوندی بین میانگین y_i و شکل خطی از پیش‌بینی‌کننده‌ها برقرار می‌کند).

ج) پیش‌بینی‌کننده

$x_i^T\beta$ پیش‌بینی کننده است که x_i^T، iامین سطر از ماتریس مدل است و β، بردار پارامترها در پیش‌بینی‌کننده است. (مک‌کلوچ و همکارش، ۲۰۰۱)

۱-۴-۴ ماتریس اطلاع فیشر

برای یک مدل شامل بردار $\boldsymbol{\beta}$ و طرح ξ که در (۱ – ۱) توصیف شده است، ماتریس اطلاع فیشر

[۱] GeneraL Linear Models

موردانتظار به صورت زیر است:

$$\mathbf{M}_{\xi}(\boldsymbol{\beta};\mathbf{x}) = E[\frac{-\partial^{2}\ell(\boldsymbol{\beta};\boldsymbol{Y})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{T}}]$$

که $\ell(\boldsymbol{\beta};\boldsymbol{Y})$ لگاریتم تابع درستنمایی مربوط به تابع جرم یا چگالی احتمال متغیر تصادفی وابسته n بعدی $\boldsymbol{Y} = (y_{1},\ldots,y_{n})^{T}$ است. برای مدل‌های خطی، ماتریس اطلاع به پارامتر $\boldsymbol{\beta}$ وابسته نیست و عبارت است از:

$$\mathbf{M}(x) = \mathbf{X}^{T}\mathbf{G}\mathbf{X}$$

است که $\mathbf{G} = diag(g_{1},\ldots,g_{n})$ و $\mathbf{X} = (\mathbf{x}_{1}^{T},\cdots,\mathbf{x}_{n}^{T})$.

در مدل‌هایی که غیرخطی‌اند، ماتریس اطلاع به پارامتر $\boldsymbol{\beta}$ وابسته است و به صورت

$$\mathbf{M}(\boldsymbol{\beta};x) = \mathbf{C}^{T}\mathbf{H}\mathbf{C}$$

محاسبه شده است، که $\mathbf{C}$ یک ماتریس $n\times p$ ای از مشتقات جزئی به شرح ذیل می‌باشد:

$$\mathbf{C} = [\frac{\partial\eta(x_{1},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}},\ldots,\frac{\partial\eta(x_{n},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}]^{T}$$

ماتریس اطلاع برای مدل‌های خطی تعمیم یافته برای داده‌های دودویی نیز به پارامترهای مدل وابسته است و به صورت

$$\mathbf{M}(\boldsymbol{\beta};x) = \mathbf{X}^{T}\mathbf{W}\mathbf{X}$$

حاصل می‌شود بطوریکه

$$\mathbf{W} = diag(\frac{w_{i}}{\pi_{i}(1-\pi_{i})}(\frac{d\pi_{i}}{d\eta_{i}})^{2}).$$

در این حالت وابستگی ماتریس اطلاع روی مقادیر پارامترها از طریق ماتریس $\mathbf{W}$ است، که عناصرش توابعی از π_{i} و x_{i} و β می‌باشند. برای مثال ماتریس اطلاع برای مدل لجستیک با داده‌های دوجمله‌ای شامل ماتریس $\mathbf{W} = diag(w_{i}\pi_{i}(1-\pi_{i}))$ است که $\pi_{i} = logit^{-1}(\mathbf{x_{i}}^{T}\boldsymbol{\beta})$.

۱-۴-۵ ماتریس اطلاع طرح مربوط به مدل‌های خطی و خصوصیات آن

ماتریس اطلاع طرح مدل‌های خطی، میانگین وزنی از ماتریس اطلاع مربوط به مشاهدات مؤلفه اول طرح است و با نماد $\mathbf{M}(\xi)$ نشان داده می‌شود

$$\mathbf{M}(\xi) = \int_{X} f(x)f^{T}(x)dx$$

یا

$$\mathbf{M}(\xi) = \sum_{i=1}^{m} w_{i}\ f(x_{i})\ f^{T}(x_{i}).$$