مجله تخصصی اپیدمیولوژی ایران؛ ۱۳۸٤، دوره ۱، شماره ۲: صفحات ٤٥-٤١.

مقاله پژوهشی

# به‌کارگیری متغیرهای پنهان در مدل رگرسیون لجستیک برای حذف اثر هم‌خطی چندگانه در تحلیل برخی عوامل مرتبط با سرطان پستان

**محمد امین پورحسینقلی[۱]، یدا... محرابی[۲]، حمید علوی مجد[۳]، پروین یاوری[٤]**

[۱]کارشناس ارشد آمار زیستی، دانشگاه علوم پزشکی شهید بهشتی، تهران.

[۲]دانشیار آمار زیستی، دانشگاه علوم پزشکی شهید بهشتی، تهران.

[۳]استادیار آمار زیستی، دانشگاه علوم پزشکی شهید بهشتی، تهران.

[۴]استاد اپیدمیولوژی، دانشگاه علوم پزشکی شهید بهشتی، تهران.

نویسنده‌ی رابط: دکتر یدا... محرابی، گروه پزشکی اجتماعی و بهداشت دانشکده پزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، اوین، کد پستی ۱۹۳۹۵، تلفن: ۸-۲۳۸۷۲۵۶۷-۰۲۱، نمابر: ۲۲۴۱۴۱۰۸-۰۲۱، پست الکترونیک: mehrabi@sbmu.ac.ir ; ymehrabi@gmail.com

تاریخ دریافت: ۸۴/۱۰/۲۶، پذیرش: ۸۵/۲/۳

**مقدمه و اهداف:** رگرسیون لجستیک یکی از کاربردی‌ترین مدل‌های خطی تعمیم‌یافته برای تحلیل رابطه‌ی یک یا چند متغیر توضیحی بر متغیر پاسخ رسته‌ای است. زمانی که بین متغیرهای توضیحی همبستگی‌های نسبتا قوی وجود داشته باشد هم‌خطی چندگانه ایجاد شده، ممکن است به کاهش کارآیی مدل منجر شود. هدف این تحقیق استفاده از متغیرهای پنهان برای کاهش اثر هم‌خطی چندگانه در تحلیل یک مطالعه مورد ـ شاهدی است.

**روش کار:** داده‌های مورد استفاده در این تحقیق متعلق به یک مطالعه مورد ـ شاهدی است که در آن ۳۰۰ نفر زن مبتلا به سرطان پستان با ۳۰۰ زن شاهد از نظر عوامل خطر مورد مقایسه قرار گرفتند. برای بررسی اثر هم‌خطی، پنج متغیر کمی که بین آن‌ها همبستگی بالایی وجود داشت، در نظر گرفته شدند. ابتدا مدل لجستیک به متغیرهای فوق برازش داده شد. سپس به منظور حذف اثر هم‌خطی، دو متغیر پنهان با استفاده از هرکدام از دو روش تحلیل عاملی و تحلیل مؤلفه‌های اصلی به‌دست آورده، بر مبنای آن‌ها پارامترهای مدل‌های لجستیک مجدداً محاسبه شدند. کارآیی مدل‌ها، با استفاده از خطای استاندارد پارامترها مقایسه گردید.

**نتایج:** مدل رگرسیون لجستیک براساس متغیرهای اولیه حاکی از مقادیر غیرعادی نسبت شانس برای سن در اولین زایمان زنده (٤٥٣٥٠٣ و ١٠١٨٤=٪CI٩٥ و OR=٦٧٩٦٠) و سن در اولین حاملگی (OR=٠/٠٠٠٠٢٩) بود. درحالی که پارامترهای مدل‌های لجستیک حاصل از متغیرهای پنهان به دست آمده از هر دو روش تحلیل عامل و تحلیل مؤلفه‌های اصلی، از نظر آماری معنی دار (p<٠/٠٠٣) و خطای استاندارد همه‌ی آن‌ها کوچک‌تر از خطای استاندارد مربوط به رگرسیون لجستیک معمولی بود. فاکتورها و مولفه‌های اصلی تولید شده توسط دو روش حداقل ۸۵ درصد کل واریانس را تبیین کردند.

**نتیجه‌گیری:** تحقیق نشان داد انحراف استاندارد پارامترهای برآورد شده در رگرسیون لجستیک براساس متغیرهای پنهان از رگرسیون لجستیک براساس مشاهدات اولیه کوچکتر بوده و در نتیجه این‌گونه مدل‌بندی در تحلیل برخی عوامل خطر سرطان پستان که هم‌خطی دارند کارآتر است.

**واژگان کلیدی:** هم‌خطی چندگانه، متغیر پنهان، تحلیل عاملی، تحلیل مؤلفه‌های اصلی، رگرسیون لجستیک، سرطان پستان.

## مقدمه

رگرسیون لجستیک یکی از کاربردی‌ترین مدل‌های خطی تعمیم‌یافته است که برای تحلیل رابطه‌ی یک یا چند متغیر توضیحی بر متغیر پاسخ رسته‌ای به کار می‌رود (۱). زمانی که تعداد متغیرهای توضیحی افزایش می یابد، مدل‌سازی مشکل شده و کارآیی آن نیز کاهش می یابد؛ به خصوص اگر برخی از متغیرها علی‌رغم فرض استقلال در مدل‌سازی، با یک‌دیگر هم‌بستگی قوی

داشته باشند و به عبارت دیگر هم‌خطی چندگانه (Multicollinearity) ایجاد شده باشد (۲). هم‌خطی چندگانه یکی از دلایل افزایش خطای استاندارد برآورد ضرایب رگرسیونی و درنتیجه کاهش کارآیی مدل بوده و ممکن است منجر به پیش‌بینی‌هایی خارج از دامنه‌ی مورد انتظار شود (۳).

مسئله‌ی هم‌خطی در مدل‌های رگرسیون خطی مورد توجه بسیاری از محققان قرار گرفته و روش‌های گوناگونی برای مقابله با اثرات نامطلوب آن ابداع شده است (۲). از جمله این روش‌ها، کاهش ابعاد مدل با استفاده از متغیرهای پنهان (Latent Variables) است. این نوع متغیرها مستقیماً مشاهده نمی‌شوند؛ بلکه از ترکیب سایر متغیرهای مشاهده‌شده قابل دستیابی بوده، به عنوان نماینده‌ی برخی از متغیرهای همبسته در مدل به‌کار می‌روند (۴).

اگر چه استفاده از متغیرهای پنهان برای کاهش ابعاد مدل، در عمل بیشترین کاربرد را در مطالعات مربوط به علوم اجتماعی و روانشناسی داشته است، ولی به دلیل نوع مطالعات انجام شده در علوم پزشکی و بهداشت که مستلزم جمع‌آوری تعداد قابل توجهی متغیرهای مرتبط با یک‌دیگر است، مشکل هم‌خطی در بسیاری از مدل‌های آماری این مطالعات قابل انتظار است (۵) و علی‌رغم این‌که هم‌خطی چندگانه در مدل رگرسیون لجستیک نیزایجاد مشکل می‌کند (۶،۷)، تاکنون توجه محققان بیشتر بر رگرسیون خطی با متغیر پاسخ دارای توزیع نرمال متمرکز بوده است.

تحلیل مؤلفه‌های اصلی (Principal Component Analysis) یکی از کاربردی‌ترین روش‌های کاهش ابعاد در روش‌های چند متغیری است. تاریخچه‌ی ابداع این روش به ابداعات پیرسن (Pearson) در برازش حداقل مربعات متعامد برمی‌گردد؛ ولی بسط عمده‌ی تئوری به‌وسیله‌ی هتلینگ (Hotteling) انجام شده است (۸). مؤلفه‌های اصلی با توجه به خصوصیاتی که دارند برای مقابله با مشکل هم‌خطی و کاهش ابعاد مدل در رگرسیون‌های خطی مورد استفاده قرار می گیرند (۲،۹). در این روش با استفاده از ماتریس مقادیر ویژه (Eigen Values) (۱۰)، مؤلفه‌های اصلی به صورت ترکیب خطی از متغیرهای اولیه و مستقل از یکدیگر ساخته می‌شوند و در آنالیز داده‌ها، به جای متغیرهای اولیه مورد استفاده قرار می‌گیرند (۱۱).

تحلیل عاملی (Factor Analysis) از دیگر روش‌های کاهش ابعاد داده‌ها است، که نخستین بار توسط اسپیرمن (Spearman) معرفی شد. در این روش با فرض وجود یک مدل مبنایی مشخص برای کل داده‌ها، و براساس ماتریس واریانس ـ کوواریانس یا ماتریس ضرایب همبستگی، عامل‌های مستقل از یکدیگر، از روی متغیرهای اولیه به‌دست می‌آیند (۱۲).

در این تحقیق، برای کاهش ابعاد مدل رگرسیون لجستیک با متغیرهای توضیحی هم‌خط، در تحلیل داده‌های یک مطالعه مورد ـ شاهدی پیرامون عوامل خطرسرطان پستان از دو روش تحلیل عاملی و تحلیل مؤلفه‌های اصلی استفاده شده است.

## روش‌ها

برای بررسی نحوه‌ی کاربرد دو روش تحلیل مؤلفه‌های اصلی و تحلیل عاملی در کاهش ابعاد مدل و ایجاد متغیرهای پنهان، از داده‌های مطالعه‌ی مورد ـ شاهدی مربوط به عوامل خطر سرطان پستان استفاده شد (۱۳). در مطالعه‌ی مذکور که در فاصله‌ی زمانی بهمن ۸۲ تا آذر ۸۳ در مرکز پزشکی ـ آموزشی، درمانی شهدای تجریش انجام شد گروه مورد، بیمارانی بودند که بیماری سرطان پستان آن‌ها با استفاده از آزمایش‌های پاتولوژیک، تشخیص قطعی داده شده و یا برای درمان یا پی‌گیری به درمانگاه بیمارستان شهدای تجریش مراجعه کرده بودند. گروه شاهد زنانی بودند که به دلایل دیگری غیر از سرطان پستان و به‌طور هم‌زمان در بخش‌های دیگرِ بیمارستان شهداء، مثل جراحی، پوست، داخلی و غیره بستری و یا برای پی‌گیری یا درمان به درمانگاه بیمارستان مراجعه کرده بودند و از نظر سنی با گروه مورد با حداکثر ۲ سال اختلاف مشابه‌سازی شدند. با اطمینان ۹۵ درصد و توان آزمون ۸۰ درصد، تعداد نمونه برای هر گروه ۳۰۰ نفر انتخاب شد (۱۴). متغیرهای مختلفی به عنوان عوامل خطر یا متغیر کنترل جمع‌آوری شدند که در این مقاله، پنج متغیر که بین آن‌ها هم‌بستگی بالایی وجود داشت در نظر گرفته شدند. این متغیرها عبارتند از: تعداد حاملگی (NP: Number of Pregnancy)، تعداد فرزندان زنده به دنیا آورده (NLB: Number of Live Birth)، کل طول مدت شیردهی به فرزندان (TLBF: Total Length of Breast Feeding)، سن در اولین حاملگی (AFP: Age at First Pregnancy) و سن در اولین زایمان زنده (AFLB: Age at First Live Birth). برای بررسی میزان بروز هم‌خطی در این مشاهدات از ماتریس ضرایب هم‌بستگی استفاده است (۲).

ابتدا بدون در نظر گرفتن وجود هم‌خطی، مدل رگرسیون لجستیک به داده‌ها برازش داده شد. سپس با ترکیب پنج متغیر مورد بررسی، یک بار با روش تحلیل مؤلفه‌های اصلی و بار دیگر به طریق تحلیل عاملی، دو متغیر پنهان به‌دست آمد و بر اساس آن‌ها، پارامترهای مدل رگرسیون لجستیک برآورد شد. مدل‌های حاصل،

براساس واریانس‌های تبیین شده به‌وسیله‌ی دو روش و خطای استاندارد پارامترهای برآورد شده، مورد مقایسه قرار گرفتند.

## یافته‌ها

جدول یک نشان می‌دهد بین سه متغیر تعداد حاملگی، تعداد فرزندان زنده به دنیا آورده و کل طول مدت شیردهی به فرزندان و نیز بین دو متغیر سن در اولین حاملگی و سن در اولین زایمان زنده هم‌بستگی خطی بالایی مشاهده می شود. هم‌چنین مقادیر عامل تورم واریانس (VIF: Variance Inflation Factor) نیز که میزان بروز هم‌خطی را نشان می دهد، محاسبه شد که همه‌ی مقادیر بیشتر از یک و نشان دهنده‌ی وجود هم‌خطی در متغیرهاست. به ویژه این‌که دو متغیر سن دراولین حاملگی و سن در اولین زایمان زنده دارای VIF بیشتر از ۱۵ بوده و بنابراین براساس این معیار هم‌خطی شدیدی دارند. آزمون بارتلت نیز نشان داد ماتریس ضرایب هم‌بستگی متغیرهای توضیحی با صفر اختلاف معنی‌داری دارد (P<۰/۰۰۱). بنابراین بین متغیرهای توضیحی مورد بررسی هم‌خطی چندگانه وجود دارد.

جدول ۱- ماتریس ضرایب همبستگی متغیرهای توضیحی

| | AFLB | AFP | TLBF | NLB | NP |
|---|---|---|---|---|---|
| NP | -۰/۴۲ | -۰/۴۴ | ۰/۶۷ | ۰/۹۰ | ۱ |
| NLB | -۰/۰۴۷ | ۰/۴۷ | ۰/۷۵ | ۱ | |
| TLBF | -۰/۴۱ | -۰/۴۲ | ۱ | | |
| AFP | ۰/۹۷ | ۱ | | | |
| AFLB | ۱ | | | | |

NP: تعداد حاملگی؛ NLB: تعداد فرزندان زنده به دنیا آورده؛ TLBF: کل طول مدت شیردهی به فرزندان؛ AFP: سن در اولین حاملگی؛ AFLB: سن در اولین زایمان زنده

نتایج حاصل از رگرسیون لجستیک بدون در نظر گرفتن وجود این هم‌خطی (جدول ۲)، نشان می‌دهد که فقط دو متغیر سن در اولین حاملگی و سن در اولین زایمان زنده معنی‌دار شده اند (P<۰,۰۰۱). هم‌چنین نسبت شانس به‌دست آمده برای سن در اولین زایمان زنده بسیاربزرگ (CI۹۵٪=۱۰۱۸۴-۴۵۳۵۰۳ و OR=۶۷۹۶۰) و برعکس برای سن در اولین حاملگی بسیار کوچک (OR=۰/۰۰۰۰۲۹) بود که هردو غیرعادی هستند. در مرحله‌ی بعد با روش تحلیل عاملی دو عامل به صورت ترکیب خطی زیر از متغیرهای اولیه بدست آمد:

Factor1= ۰/۸۵NP+۰/۹۷NLB+۰/۶۸TLBF-۰/۲۶AFP-۰/۲۶AFLB

Factor2= ۰/۲۱NP-۰/۲۳NLB-۰/۲۵TLBF+۰/۹۴AFP+۰/۹۵AFLB

در مجموع ۸۴/۷۹٪ واریانس توسط عامل اول (۴۵/۶۷٪) و عامل دوم (۳۹/۱۲٪) تبیین شده است. وارد کردن عوامل فوق به عنوان متغیرهای توضیحی در مدل رگرسیون لجستیک، نشان داد که عامل اول (P<۰/۰۰۲) و عامل دوم (P<۰/۰۰۱) هر دو تاثیر معنی‌داری بر متغیر وابسته دارند (جدول ۳). در مرحله‌ی سوم، با روش تحلیل مؤلفه‌های اصلی دو مؤلفه به صورت ترکیب خطی زیر از متغیرهای اولیه بدست آمد:

Component1=۰/۹۰NP+۰/۹۲NLB+۰/۸۲TLBF-۰/۲۵AFP

-۰/۲۴AFLB

Component2=-۰/۲۱NP-۰/۲۵NLP-۰/۲۵TLBF+۰/۹۶AFP

+۰/۹۶AFLB

مؤلفه‌ی اول ۴۹/۲۶٪ و مؤلفه‌ی دوم ۴۰٪ و در مجموع ۸۹/۲۶٪ کل واریانس را تبیین می‌کنند. در برآورد پارامترهای رگرسیون لجستیک براساس این مؤلفه‌های اصلی، نتایج مشابهی حاصل شد. به عبارت دیگر مؤلفه اول با (P<۰/۰۰۲) و مؤلفه دوم با (P<۰/۰۰۳) معنی‌دار شدند و نسبت شانس‌ها نیز تقریباً مشابه

جدول ۲- برآورد پارامترهای مدل رگرسیون لجستیک بدون درنظر گرفتن همخطی بین متغیرهای توضیحی

| متغیرهای اصلی | ضرایب رگرسیونی | خطای استاندارد | P-value | نسبت شانس (فاصله اطمینان ۹۵٪) | عامل تورم واریانس (VIF)* |
|---|---|---|---|---|---|
| عرض از مبداء | ۰/۱۰ | ۰/۲۰ | ۰/۶۲ | (۱۰۱۸۴-۴۵۳۵۰۳) | |
| تعداد حاملگی (NP) | ۰/۰۶ | ۰/۰۹ | ۰/۵۰ | ۱/۰۶ (۰/۸۹-۱/۲۷) | ۴/۵۷ |
| تعداد فرزندان زنده به دنیا آورده (NLB) | -۰/۱۷ | ۰/۲۶ | ۰/۵۱ | ۰/۸۴ (۰/۵۰-۱/۴۰) | ۵/۸۳ |
| کل طول مدت شیردهی به فرزندان (TLBF) | -۰/۲۲ | ۰/۱۵ | ۰/۱۴ | ۰/۸۰ (۰/۵۹-۱/۰۸) | ۲/۱۳ |
| سن در اولین حاملگی (AFP) | -۱۰/۴۶ | ۰/۹۲ | <۰/۰۰۱ | ۰/۰۰۰۰۲۹ | ۱۵/۴۱ |
| سن در اولین زایمان زنده (AFLB) | ۱۱/۱۳ | ۰/۹۷ | <۰/۰۰۱ | ۶۷۹۶۰ (۱۰۱۸۴-۴۵۳۵۰۳) | ۱۵/۳۴ |

*VIF: Variance Inflation Factor

جدول ۳- برآورد پارامترهای رگرسیون لجستیک بر اساس متغیرهای پنهان ایجاد شده به‌وسیله روش تحلیل عاملی و مؤلفه‌های اصلی

| متغیرهای پنهان | ضرایب رگرسیونی | خطای استاندارد | P-value | نسبت شانس (فاصله اطمینان ۹۵٪) |
|---|---|---|---|---|
| روش تحلیل عاملی | | | | |
| عرض از مبداء | -۰/۰٤ | ۰/۰۸ | ۰/٦٤ | ۰/۹٦ |
| عامل ۱ | -۰/۲۷ | ۰/۰۹ | ۰/۰۰۲ | ۰/۷٦ (۰/٦٤-۰/۹۱) |
| عامل ۲ | ۰/۲۹ | ۰/۰۹ | ۰/۰۰۱ | ۱/۳۳ (۱/۱۲-۱/۵۹) |
| روش تحلیل مؤلفه های اصلی | | | | |
| عرض از مبداء | -۰/۰٤ | ۰/۰۹ | ۰/٦٤ | ۰/۹٦ |
| مؤلفه ۱ | -۰/۲٦ | ۰/۰۹ | ۰/۰۰۲ | ۰/۷۷ (۰/٦۵-۰/۹۱) |
| مؤلفه ۲ | ۰/۲۷ | ۰/۰۹ | ۰/۰۰۳ | ۱/۳۱ (۱/۱۰-۱/۵٦) |

روش تحلیل عاملی به‌دست آمدند (جدول ۳).

## بحث

هدف پژوهش حاضر استفاده از متغیرهای پنهان در مدل رگرسیون لجستیک به منظور کاهش اثر در حالت بروز هم‌خطی چندگانه بود. نتایج حاصل نشان داد برآورد پارامترهای مدل در هر دو روش تحلیل مؤلفه‌های اصلی و تحلیل عامل، مشابه هستند و خطای استاندارد آن‌ها نسبت به مدل اصلی بسیار کوچک‌تر است. بنابراین دو روش به‌کار رفته برای تولید متغیرهای پنهان نسبت به مدل رگرسیون لجستیک براساس مشاهدات اولیه، در تحلیل برخی عوامل مرتبط با سرطان پستان ازکارایی بالاتری برخوردار هستند. در مدل اولیه به دلیل وجود هم‌خطی و بالا بودن خطای استاندارد برآوردها، تعدادی از متغیرها معنی‌دار نشده‌اند و از بین دو متغیر معنی‌دار، متغیر سن در اولین زایمان زنده، نسبت شانس بسیار بزرگ و غیر معمول و متغیر سن در اولین حاملگی، نسبت شانسی نزدیک به صفر را نشان داد.

یافته‌های این تحقیق با نظر آگویلرا و اسکابیاس (Aguilera & Escabias) که در مقاله‌ی خود نشان داده‌اند استفاده از تحلیل مؤلفه‌های اصلی در رگرسیون لجستیک با داده‌های هم‌خط می‌تواند برآورد پارامترها را بهبود بخشد، مطابق است (۱۵). این روش در مقاله‌ی کاربردی اسکابیاس، آگویلرا و والدراما (Valderrama) که در مورد مدل‌سازی داده‌های هواشناسی بود نیز مورد استفاده قرار گرفته است. آنان در این مطالعه بر نحوه‌ی انتخاب مؤلفه‌های اصلی برای بهبود برآورد پارامترها تأکید کرده‌اند (۱۶).

وال و لی (Wall & Li) در مطالعه ای که براساس متغیرهای توصیفی میزان مرگ و میر بیماری‌های تنفسی بود، نتایج حاصل از دو روش تحلیل عامل و معادله‌ی مدل‌سازی ساختاری (Structural Equation Modeling) را با رگرسیون کلاسیک براساس متغیرهای اولیه هم‌خط مقایسه کرده نشان دادند که متغیرهای پنهان، پارامترهایی با خطاهای استاندارد کوچکتر تولید می‌کنند (۱۷). نتایج تحقیق حاضر از این لحاظ با مطالعه‌ی آنان هم‌خوانی دارد.

ایده‌ی استفاده از متغیرهای پنهان به‌جای متغیرهای اصلی، با هدف کاهش ابعاد داده‌ها از این حقیقت ناشی می‌شود که این متغیرها می‌توانند بازتاب‌دهنده‌ی ارتباط بین مشاهدات باشند (۱۸). با این حال استفاده از مدل‌های دربرگیرنده‌ی متغیرهای پنهان تبعاً مزایا و محدودیت‌هایی دارد. یکی از اهداف اصلی در ساختن مدل‌های آماری تفسیر مدل با توجه به پارامترهای برآوردشده می‌باشد؛ ولی تفسیر مدل‌هایی که براساس عامل‌ها یا تحلیل مؤلفه‌های اصلی به‌دست می‌آیند قدری پیچیده است (۱۱،۱۹). برای این کار استفاده از روش تحلیل مؤلفه‌های اصلی بهتر از تحلیل عاملی است؛ زیرا مؤلفه‌های اصلی صرفاً ترکیبی خطی از متغیرهای اولیه هستند و بر خلاف روش تحلیل عامل، مدلی برای داده‌ها فرض نمی‌کند (۱۱،۲۰،۱۹). در نتیجه از طریق معکوس ماتریس دوران می‌توان برآوردهای تصحیح شده پارامترهای متغیرهای اولیه را به‌دست آورد (۱۱). هم‌چنین روش‌هایی نیز برای تفسیر این مؤلفه‌ها در مدل کاهش‌یافته پیش‌نهاد شده است (۲۱). به هر حال سودمندی‌های حاصل از کاهش ابعاد مدل و کاستن تعداد متغیرها آن‌چنان قابل ملاحظه است که علی‌رغم مشکلات حاصل در تفسیر پارامتر، برخی تکنیک‌های جدید علاوه بر تولید متغیرهای پنهان برای متغیرهای توضیحی، اکنون بر تولید این متغیرها برای متغیرهای پاسخ توجه دارند (۲۲).

## نتیجه‌گیری

براساس یافته‌های این تحقیق می‌توان نتیجه‌گیری کرد که در بررسی برخی عوامل خطر سرطان پستان، دو روش تحلیل عاملی و تحلیل مؤلفه‌های اصلی نتایج مشابهی داشته، نسبت به مدل لجستیک با متغیرهای هم‌خط اولیه از کارآیی بالاتری برخوردار هستند.

## تشکر و قدردانی

در این مقاله از داده‌های طرح تحقیقاتی عوامل خطر سرطان

پستان، مصوب دانشگاه علوم پزشکی شهید بهشتی استفاده شده است که به این وسیله از کلیه همکاران طرح مذکور و نیز از معاونت پژوهشی دانشکده پزشکی سپاس‌گزاری به عمل می آید.

## منابع


1. Myers R.H., Montgomery D.C. and Vining G.G., Generalized linear models with application in engineering and sciences, 2002, John Wiley & Sons.
2. Chattergee, S., Hadi, A.S. and Price, B. (2000). Regression analysis by example, 2002, John Wiley & Sons, USA. PP: 225-258.
3. Myers, R.H. (1990). Classical and modern regression with applications., 1990, Pws-Kent publishing company. PP: 123-129.
4. Van Eye, A., Clogg, C.C., Latent variables analysis; application for developing research. 1994, SAGE publication. PP: 3-35.
5. Hazard munro, B. Statistical methods for health care research., 2001, Philadelphia: Lippincott. PP: 287-288.
6. Kleinbaum, D. Logistic Regression., 1994, Springer, New York. PP: 168.
7. Hosmer, D.W., Lemeshow, S. Applied logistic regression., 1989, John Wiley & Sons.
8. Morrison, D. F. Multivariate statistical methods. 2002, John Wiley & Sons. PP: 312-398.
9. Rawlings, J. O. Applied regression analysis: A research tools., 1988, Belmont: Wadsworth. PP: 327-356.
10. Schott, J. R. Matrix analysis for statistics., 1997, John Wiley & Sons. PP: 84-131.
11. Jollife, I.T. PrinCI95%pal component analysis., 1986, Springer. PP: 129-141.
12. Srivastava, M. S. Methods of multivariate statistics, 2002, John Wiley & Sons. New York. PP: 397-450.
13. Yavari, P., Mousavizadeh, M., Sadrol-Hafezi, B. and Mehrabi, Y., Reproductive characteristics and the risk of breast cancer, A case-control study. Asian PaCI95%fic J Cancer Prev, 2005, 6, 370-375.
14. Lemeshow, S., Hosmer, D. W. and Klar, J. Adequacy of sample size in Health studies. World Health Organization, 1998, John Wiley & Sons. PP: 19.
15. Aguilera, A.M. and Escabias, M., PrinCI95%pal component logistic regression. Proceedings in computational statistics, 2000, 175-180. Physica-Verlag.
16. Escabias, M., Aguilera, A. M. and Valderrama, M. J., Modeling climatological data by functional logistic regression. The ISI International Conference on Environmental Statistics and Health, 2003.
17. Wall. M. M. and Li, R., Comparison of multiple regression to two latent variable techniques for estimation and prediction. Statistics in Medicine; 2003, 22:3671-3685.
18. Sobel, M. E. Causal inference in latent variable models. In Latent variables analysis; application for developing research. By Van Eye, A., Clogg, 1994, C.C SAGE publication. PP: 3-35.
19. Rencher, A. C. Methods of multivariate analysis, 2002, John Wiley & Sons.
20. Armitage, P. and Colton, T., Encyclopedia of Biostatistics. Volume 2. Chichester: 1998, John Wiley & Sons. PP: 1480-1481.
21. Chipman HA and Gu H. Interpretable dimension reduction. 2002, http://ace.acadiau.ca/math/chipmanh/publications.html.
22. Guo, J., Wall, M. M. and Amemiya Y. Latent class regression on latent factors to appear in Biostatistics.