بسم الله الرحمن الرحيم
الحمد لله رب العالمين الرحمن الرحيم مالك يوم الدين
اهدنا الصراط المستقيم صراط الذين
انعمت عليهم غير المغضوب عليهم
ولا الضالين

بسمه تعالی

دانشگاه تربیت مدرس

دانشکده علوم ریاضی

تأییدیه اعضای هیأت داوران حاضر در جلسه دفاع از پایان‌نامه کارشناسی‌ارشد

اعضای هیأت داوران نسخهٔ نهایی پایان‌نامه خانم کبری قلی‌زاده‌گزور رشته آمار به شماره دانشجویی ۸۹۵۲۰۲۱۰۰۶ تحت عنوان: «تحلیل بیزی مدل‌های رگرسیون جمعی ساختاری با استفاده از تقریب لاپلاس آشیانی جمع‌بسته» را در تاریخ ۱۳۹۱/۱۰/۲۰ از نظر فرم و محتوا بررسی نموده و آن را برای اخذ درجهٔ کارشناسی‌ارشد مورد تأیید قرار دادند.

اعضای هیأت داوران	نام و نام‌خانوادگی	رتبه علمی	امضاء
۱- استاد راهنما	دکتر محسن محمدزاده	استاد	
۲- استاد ناظر داخلی	دکتر مجید جعفری‌خالدی	استادیار	
۳- استاد ناظر داخلی	دکتر موسی گل‌علی‌زاده	استادیار	
۴- استاد ناظر خارجی	دکتر حمید پزشک	استاد	
۵- نماینده تحصیلات تکمیلی	دکتر موسی گل‌علی‌زاده	استادیار	

آیین نامه چاپ پایان‌نامه (رساله)های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله)های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است بنابراین به منظور آگاهی و رعایت حقوق دانشگاه‌دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می‌شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله)ی خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثارعلمی» دانشگاه اطلاع دهد.

ماده ۲: در صفحه سوم کتاب (پس از برگ شناسنامه) عبارت ذیل را چاپ کند:

«کتاب حاضر، حاصل پایان نامه کارشناسی ارشد/ رساله دکتری نگارنده در رشته آمار است که در سال ۱۳۹۱ در دانشکده علوم ریاضی دانشگاه تربیت مدرس به راهنمایی ~~سرکار خانم~~/جناب آقای دکتر محمد زاره، مشاوره سرکار خانم/جناب آقای دکتر ——— و مشاوره سرکار خانم/جناب آقای دکتر ——— از آن دفاع شده است.»

ماده ۳: به منظور جبران بخشی از هزینه‌های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت‌چاپ) را به «دفتر نشر آثارعلمی» دانشگاه اهدا کند. دانشگاه می‌تواند مازاد نیاز خود را به نفع مرکز نشر درمعرض فروش قرار دهد.

ماده ۴: در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده را به عنوان خسارت به دانشگاه تربیت‌مدرس، تأدیه کند.

ماده ۵: دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت بهای خسارت، دانشگاه می‌تواند خسارت مذکور را از طریق مراجع قضایی مطالبه و وصول کند؛ به علاوه به دانشگاه حق می‌دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶: اینجانب کسری قلی زاده گزور دانشجوی رشته آمار مقطع کارشناسی ارشد تعهد فوق وضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی: کسری قلی زاده گزور

تاریخ و امضا:
۹۱،۱۰،۲۷

آیین‌نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی دانشگاه تربیت مدرس

مقدمه: با عنایت به سیاست‌های پژوهشی و فناوری دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیأت علمی، دانشجویان، دانش‌آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان‌نامه، رساله و طرحهای تحقیقاتی با هماهنگی دانشگاه انجام شده است، موارد زیر را رعایت نمایند:

ماده ۱- حق نشر و تکثیر پایان نامه/ رساله و درآمدهای حاصل از آنها متعلق به دانشگاه می باشد ولی حقوق معنوی پدید آورندگان محفوظ خواهد بود.

ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان‌نامه/ رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و با تایید استاد راهنمای اصلی، یکی از اساتید راهنما، مشاور و یا دانشجو مسئول مکاتبات مقاله باشد. ولی مسئولیت علمی مقاله مستخرج از پایان نامه و رساله به عهده اساتید راهنما و دانشجو می باشد.

تبصره: در مقالاتی که پس از دانش‌آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان‌نامه/ رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

ماده ۳- انتشار کتاب، نرم افزار و یا آثار ویژه (اثری هنری مانند فیلم، عکس، نقاشی و نمایشنامه) حاصل از نتایج پایان‌نامه/ رساله و تمامی طرحهای تحقیقاتی کلیه واحدهای دانشگاه اعم از دانشکده ها، مراکز تحقیقاتی، پژوهشکده ها، پارک علم و فناوری و دیگر واحدها باید با مجوز کتبی صادره از معاونت پژوهشی دانشگاه و براساس آئین نامه های مصوب انجام شود.

ماده ٤- ثبت اختراع و تدوین دانش فنی و یا ارائه یافته ها در جشنوارههای ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان‌نامه/ رساله و تمامی طرح‌های تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق معاونت پژوهشی دانشگاه انجام گیرد.

ماده ٥- این آیین‌نامه در ٥ ماده و یک تبصره در تاریخ۸۷/٤/۱ در شورای پژوهشی و در تاریخ ۸۷/٤/۲۳ در هیأت رئیسه دانشگاه به تایید رسید و در جلسه مورخ ۸۷/۷/۱۵ شورای دانشگاه به تصویب رسیده و از تاریخ تصویب در شورای دانشگاه لازم‌الاجرا است.

«اینجانب کبری علی زاده کنور دانشجوی رشته آمار ورودی سال تحصیلی ۱۳۸۹
مقطع کارشناسی ارشد دانشکده علوم ریاضی متعهد می شوم کلیه نکات مندرج در آئین نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهش های علمی دانشگاه تربیت مدرس را در انتشار یافته های علمی مستخرج از پایان نامه / رساله تحصیلی خود رعایت نمایم. در صورت تخلف از مفاد آئین نامه فوق الاشعار به دانشگاه وکالت و نمایندگی می دهم که از طرف اینجانب نسبت به لغو امتیاز اختراع ثبت نام ننده و یا هر گونه امتیاز دیگر و تغییر آن به نام دانشگاه اقدام نماید. ضمناً نسبت به جبران فوری ضرر و زیان حاصله بر اساس برآورد دانشگاه اقدام خواهم نمود و بدینوسیله حق هر گونه اعتراض را از خود سلب نمودم»

امضا.....

تاریخ ۲۷/۱۰/۹۱

دانشگاه تربیت مدرس

دانشکده علوم ریاضی

پایان نامه دوره کارشناسی ارشد آمار

تحلیل بیزی مدل‌های رگرسیون جمعی ساختاری با استفاده از تقریب لاپلاس آشیانی جمع‌بسته

توسط

کبری قلی‌زاده گزور

استاد راهنما

دکتر محسن محمدزاده

دی ماه ۱۳۹۱

اکنون که برگذری دیگر از زندگی ایستاده‌ام حاصل تلاش و اشتیاق روزهای متمادی‌ام را تقدیم می‌کنم به

«او که روزی خواهد آمد»

و تقدیم به پدر و مادر مهربانم.

قدردانی

آن‌که مرا کلامی می آموزد مرا بنده خود می‌کند.

شکر و سپاس ایزد دانا را که توان آموختن عطا فرمود و این یگانه معلم هستی، معلمان را برترین جانشینان خود در عرصه گیتی برگزید. اکنون که به مدد ایزد توانا توانسته‌ام پایان‌نامه کارشناسی ارشد خود را فراهم آورم بر خود لازم می‌دانم کمال تقدیر و تشکر را از دکتر محسن محمدزاده که راهنمای این پایان‌نامه را بر عهده داشته و مرا از راهنمایی‌های بی‌دریغشان بهره‌مند نموده‌اند داشته باشم و از یگانه دوست برای ایشان و خانواده محترمشان توفیق و سعادت روز افزون خواستارم.

اکنون که در حال نگارش این سطور هستم خاطراتم آرام آرام ورق می‌خورند، روزهایی که مادرم چه کودکانه زانو بر زمین زد تا بلندای حضورش را به اندازه همه کودکی من، به کودکی غرق در غرور کودکانه‌اش هدیه کند و او را مملو از عشق مادرانه رهسپار سرزمین مقدس دانستن کند. پدرم که همواره تکیه بر استواری قامتش زده و چون نیلوفری سعی در قد کشیدن کرده‌ام و نیک می‌دانم اگر استواری قامت همچو سروش نبود تمام تلاش من هرگز راهی برای دیدن آسمان نمی‌یافت.

این تحقیق با حمایت پژوهشکده آمار انجام شده است، لذا بر خود لازم می‌دانم تا از ایشان قدردانی نمایم.

کبری قلی‌زاده گزور

دی ۱۳۹۱

چکیده

مدل‌های رگرسیون جمعی ساختاری قالبی انعطاف پذیر از مدل‌های آماری در زمینه‌های کاربردی هستند. گاهی در تحلیل بیز سلسله‌مراتبی این مدل‌ها توزیع‌های پسینی فرم بسته‌ای ندارد و استفاده از الگوریتم‌های مونت کارلوی زنجیره مارکوفی (MCMC) ممکن است به دلیل پیچیده بودن و تعداد زیاد پارامترهای این مدل زمان‌بر باشد. برای حل این مشکل می‌توان از تقریب لاپلاس آشیانی جمع بسته استفاده کرد، که در آن با استفاده از تقریب‌های گاوسی و لاپلاس نیاز به شبیه‌سازی‌های سنگین MCMC نیست.

در این پایان‌نامه ضمن مطالعهٔ مدل‌های رگرسیونی جمعی ساختاری، میدان تصادفی گاوسی مارکوفی که در این مدل‌ها مورد استفاده قرار می‌گیرد مطالعه می‌شود. سپس مدل‌های رگرسیونی جمعی ساختاری و نحوهٔ برآورد آن با استفاده از روش تقریب لاپلاس آشیانی جمع‌بسته برای داده‌های جرم شهر تهران بیان می‌شود و مدل برتر بر اساس ملاک‌های DIC و نمرهٔ لگاریتمی اعتبار سنجی متقابل انتخاب و تحلیل می‌شود. سپس با مطالعه‌ای شبیه‌سازی، دقت و زمان محاسبهٔ مدل‌ها با استفاده از دو روش تقریب لاپلاس آشیانی جمع بسته و MCMC مورد مقایسه قرار می‌گیرند. در انتها نتایج به‌دست آمده مورد بحث و نتیجه گیری قرار می‌گیرد.

واژه‌های کلیدی : مدل رگرسیون جمعی ساختاری، میدان تصادفی مارکوفی گاوسی، تقریب لاپلاس آشیانی جمع‌بسته، MCMC.

فهرست مندرجات

۱ مبانی و مفاهیم مقدماتی ۱

۱.۱ مقدمه ۱

۲.۱ تعاریف و مفاهیم اولیه ۳

۱.۲.۱ همبستگی فضایی ۵

۲.۲.۱ گراف ٦

۳.۲.۱ ماتریس همیشه مثبت متقارن ۹

۳.۱ مدل رگرسیونی جمعی ساختاری ۱۰

۲ مقدمه‌ای بر میدان‌های تصادفی مارکوفی گاوسی ۱۵

۱.۲ مقدمه ۱۵

۲.۲ میدان تصادفی مارکوفی گاوسی ۱۹

۳.۲ روش‌های عددی برای ماتریس‌های تُنک ۲۴

۱.۳.۲ تجزیهٔ یک ماتریس تُنک ۲۵

۲.۳.۲ مرتب‌سازی ۳۱

۴.۲ میدان تصادفی مارکوفی گاوسی ناسره ۳۳

۵.۲ میدان‌های تصادفی مارکوفی گاوسی ذاتی مراتب بالا ۳۸

۶.۲ مدل بسیج ۴۵

۷.۲ مدل قدم زدن تصادفی زمان پیوسته ۴۶

۳ تقریب لاپلاس آشیانی جمع‌بسته ۵۰

۱.۳ تقریب لاپلاس ۵۱

۱.۱.۳ میانگین و واریانس پسینی ۵۲

۲.۱.۳ چگالی پسینی کناری ۵۴

۲.۳ تقریب لاپلاس آشیانی جمع‌بسته ۵۴
۱.۲.۳ تقریب پسینی توأم بردار ابرپارامتر با راهبرد طرح مرکب مرکزی . . . ۵۵
۲.۲.۳ تقریب پسینی کناری درایه‌های بردار ابرپارامتر ۶۳
۳.۲.۳ تقریب پسینی کناری عناصر میدان پنهان ۶۳

۳.۳ ملاک‌های ارزیابی ۷۰
۱.۳.۳ تقریب کناری درستنمایی ۷۰
۲.۳.۳ ملاک انحراف اطلاع ۷۱
۳.۳.۳ ملاک پیشگویی شرطی مؤلفه‌ها ۷۳

۴ مدل‌بندی انواع جرم در شهر تهران ۷۴

۱.۴ معرفی متغیرها ۷۷

۲.۴ مدل‌بندی عوامل وقوع جرم ۸۰

۳.۴ مدل‌بندی نزاع و درگیری ۸۱
۱.۳.۴ تحلیل کاوشگرانه ۸۲

۲.۳.۴ مدل‌بندی ۸۳

۴.۴ مدل‌بندی سرقت ۹۳
۱.۴.۴ تحلیل کاوشگرانه ۹۴
۲.۴.۴ مدل‌بندی ۹۴

۵.۴ مطالعهٔ شبیه‌سازی ۹۹

۶.۴ بحث و نتیجه‌گیری ۱۰۲

لیست اشکال

۱.۲.۱ نمایی از یک گراف ساده ۷

۲.۲.۱ نقشه تهران به‌همراه گراف متناظر با آن ۸

۱.۲.۲ (الف) ویژگی مارکوفی دوبه‌دو. (ب) ویژگی مارکوفی موضعی. (ج) ویژگی مارکوفی فراموضعی ۲۲

۲.۳.۲ گراف ساده ۲۷

۳.۳.۲ (الف) گراف قبل از مرتب‌سازی. (ب) گراف بعد از مرتب‌سازی ۳۱

۴.۴.۲ نمونه‌های تولید شده از میدان تصادفی مارکوفی گاوسی ناسره ۳۸

۱.۲.۳ موقعیت نقاط در طرح مرکب مرکزی ٦۲

۲.۲.۳ نمایش ماتریسی معادله تاکاهاشی ٦۵

۱.۳.۴ پهنه‌بندی اثرات فضایی مناطق ۲۲گانه شهر تهران ۸٦

۲.۳.۴ پهنه‌بندی اثرات تصادفی مناطق ۲۲گانه شهر تهران ۸٦

۳.۳.۴ پهنه‌بندی میزان طلاق مناطق ۲۲گانه شهر تهران ۸۷

۴.۳.۴ تاثیر طلاق بر میزان نزاع و درگیری ۸۸

۵.۳.۴ سرانه فضای سبز ۹۰

٦.۴.۴ برآورد اثر جمعیت شناور ۹۷

فصل ۱

مبانی و مفاهیم مقدماتی

۱.۱ مقدمه

دگرگونی‌های عمده در حجم و پیچیدگی داده‌ها در کنار پیشرفت‌های اساسی روش‌های آماری و ارتقای تکنولوژی فناوری اطلاعات ما را به سمت استفاده از مدل‌های پیچیده اما دقیق‌تر سوق می‌دهد. معمولاً مدل‌های خطی[۱] (LM) با فرض پیوسته بودن متغیر پاسخ به کار می‌روند اما در عمل ممکن است متغیر پاسخ پیوسته نباشد به همین دلیل مدل خطی تعمیم‌یافته[۲] (GLM) توسط نلدر و ودرنبرن (۱۹۷۲) معرفی شد. در مدل خطی و GLM فرض بر این است که متغیرهای تبیینی اثر خطی بر متغیر پاسخ دارند اما در بعضی مسائل ممکن است متغیر تبیینی اثر غیر خطی داشته باشند. به منظور در نظر گرفتن توابعی هموار از این متغیرها، مدل جمعی تعمیم‌یافته[۳] (GAM)(هیستی و تیبشرانی، ۱۹۹۰) معرفی شد. در LM، GLM و GAM فرض بر این است که متغیرهای پاسخ مستقل هستند، اما گاهی در

[۱] Linear Models

[۲] Genralized Linear Model

[۳] Generalized Additive Model

عمل مواردی وجود دارند که متغیرهای پاسخ وابسته‌اند. بریسلو و کلیتون (۱۹۹۳) با اضافه کردن اثرات تصادفی و پذیرش فرض استقلال شرطی متغیر پاسخ در GLM مدل آمیخته خطی تعمیم‌یافته [۴] (GLMM) را معرفی کردند. علاوه بر این لین و ژانگ (۱۹۹۹) مدل آمیخته جمعی تعمیم‌یافته [۵] (GAMM) را به‌عنوان تعمیمی از GAM برای داده‌هایی با متغیر پاسخ وابسته معرفی کردند، به‌طوری که این وابستگی از طریق یک متغیر تبیینی در مدل لحاظ می‌شود. فهرامیر و تاتز (۲۰۰۱) مدل کلی رگرسیون جمعی ساختاری [۶] (STAR) را معرفی کردند، که در آن متغیر پاسخ متعلق به خانواده نمایی است و متغیرهایی تبیینی با اثرات خطی و غیر خطی نیز در مدل لحاظ می‌شوند. به‌علاوه ناهمگونی متغیرهای کمکی می‌تواند در مدل مورد توجه قرار گیرد و شرط استقلال شرطی متغیر پاسخ جایگزین فرض استقلال شود. اغلب مدل‌های ذکر شده زیر کلاسی از مدل‌های رگرسیون جمعی ساختاری هستند.

برای تحلیل این مدل‌ها با رهیافت بیزی نیاز به محاسبهٔ توزیع‌های پسینی است، که معمولاً الگوریتم‌های مونت کارلوی زنجیره مارکوفی [۷] (MCMC) می‌تواند روشی مفید در به‌دست آوردن آن‌ها و تحلیل بیزی مدل‌ها باشد. اما در عمل این الگوریتم‌ها به دلیل پیچیده بودن مدل‌های STAR ممکن است با مشکلی همچون طولانی بودن زمان محاسبات مواجه شود که برای حل این مشکل رو و همکاران (۲۰۰۹) روش تقریب لاپلاس آشیانی جمع بسته [۸] (INLA) را معرفی کردند و نشان دادند این روش ضمن حفظ دقت برآورد پارامترها سرعت محاسبات را نیز افزایش می‌دهد. فانگ و همکاران (۲۰۱۰) نشان دادند روش INLA برای مدل‌های GLMM در حالت کلی دقیق است، اما

[۴] Generalized Linear Mixed Model
[۵] Generalized Additive Mixed Model
[۶] Structured Additive Regression Model
[۷] Markov Chain Monte Carlo
[۸] Integrated Nested Laplace Approximation

برای داده‌های دوجمله‌ای با تعداد آزمایشات کم، دقت کمتری دارد. رز و هلد (۲۰۱۱) با به کاربردن روش INLA به بررسی حساسیت مدل‌های GLMM به پیشینی‌های فرض شدهٔ ابرپارمترها پرداختند و بر اساس فاصله هلینگر که شباهت بین دو توزیع احتمال را می‌سنجد، اندازه حساسیت را توسعه دادند، همچنین برای انتخاب مدل چندین تکنیک اعتبار سنجی متقابل پیشنهاد دادند. اخیراًاسکلر و هلد (۲۰۱۱) چگونگی استفاده از INLA برای استنباط بیزی انواع مدل‌های فضایی– زمانی را نشان دادند و برای داده‌های اسهال ویروسی گاو در سوئیس دو روش INLA و MCMC را از نظر دقت با هم مقایسه کردند. همچنین قیومی و همکاران (۱۳۹۱) نحوهٔ کاربست INLA را در مدل‌های گاوسی پنهان فضایی بر روی داده‌های میزان ضعف بدنی کودکان در کشور زامبیا به نمایش گذاشته و دقت نتایج و سرعت محاسبات این روش و الگوریتم‌های MCMC را مورد مقایسه و ارزیابی قرار دادند.

در ادامهٔ این فصل مفاهیم و مقدمات اولیه و مدل رگرسیون جمعی ساختاری بیان می‌شود. در فصل ۲ میدان تصادفی گاوسی، روش‌های عددی برای ماتریس‌های تنک و انواع آن تعریف می‌شوند. در فصل ۳ با تعریف روش لاپلاس به معرفی روش تقریب لاپلاس آشیانی جمع بسته پرداخته می‌شود. در فصل ۴ نحوهٔ کاربست روش‌های ارائه شده در فصل‌های قبل برای مدل‌بندی و تحلیل داده‌های جرم مناطق ۲۲گانه شهر تهران ارائه شده و در انتها به بحث و نتیجه‌گیری پرداخته می‌شود.

۲.۱ تعاریف و مفاهیم اولیه

در این پایان‌نامه بردارها و ماتریس‌ها به‌صورت پُررنگ، مانند $\boldsymbol{x}$ و $\boldsymbol{A}$ نشان داده می‌شوند. ترانهادهٔ $\boldsymbol{A}$ با $\boldsymbol{A}^T$ مشخص می‌شود. نماد $\boldsymbol{A} = (A_{ij})$ مشخص می‌کند که مولفهٔ سطر iام و ستون jام ماتریس $\boldsymbol{A}$، A_{ij} است. همچنین مشابه این نماد نیز برای بردار استفاده می‌شود، به‌عبارتی $\boldsymbol{x} = (x_i)$. $\boldsymbol{x}_{i:j}$

کوتاه شدهٔ بردار $\boldsymbol{x}$ به‌صورت $(x_i, x_{i+1}, \ldots, x_j)^T$ است. اگر $\boldsymbol{A}_{n\times m}$ ماتریسی با m ستون $\boldsymbol{A}_m, \ldots, \boldsymbol{A}_1$ باشد، $Vec(\boldsymbol{A}) = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m)$ برداری ستونی است که از زیر هم قرار دادن ستون‌های ماتریس $\boldsymbol{A}$ حاصل می‌شود. ماتریسی که از حذف بعضی از سطرها و ستون‌های $\boldsymbol{A}$ به‌دست می‌آید زیرماتریس $\boldsymbol{A}$ نامیده می‌شود. زیر ماتریسی را که از حذف سطر و ستون با اندیس یکسان به‌دست آید زیرماتریس اصلی گویند. به‌عنوان مثال، ماتریس $\boldsymbol{B} = \begin{pmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{pmatrix}$ که از حذف سطر و ستون دوم ماتریس $\boldsymbol{A}$ حاصل شده است یک زیرماتریس اصلی ماتریس $\boldsymbol{A}$ است. زیر ماتریس $r \times r$، که از حذف $n-r$ سطر و ستون آخر ماتریس $\boldsymbol{A}$ به‌دست می‌آید زیر ماتریس اصلی مقدم[۹] $\boldsymbol{A}$ نامیده می‌شود.

اگر $\boldsymbol{A}$ ماتریس $n \times n$ و $\boldsymbol{a}$ برداری $n \times 1$ باشد، آن‌گاه $diag(\boldsymbol{A})$ و $diag(\boldsymbol{a})$ به‌ترتیب به‌منظور نمایش ماتریس‌های قطری

$$diag(\boldsymbol{A}) = \begin{pmatrix} A_{11} & & \\ & \ddots & \\ & & A_{nn} \end{pmatrix}, \qquad diag(\boldsymbol{a}) = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$$

استفاده می‌شوند. پهنای‌باند ماتریس $\boldsymbol{A}$ به‌صورت $\max\{|i-j| : A_{ij} \neq 0\}$ تعریف می‌شود. به‌طور مشابه پهنای‌باند پایینی $\boldsymbol{A}$ برابر با $\max\{|i-j| : A_{ij} \neq 0, \quad i > j\}$ و پهنای باند بالایی برابر $\max\{|i-j| : A_{ij} \neq 0, \quad i < j\}$ است. مجموعهٔ $\mathcal{I}_{\boldsymbol{n}} = \{(i,j) : i = 1, \ldots, n_1, \quad j = 1, \ldots, n_2\}$ نشانگر مشبکهٔ منظم[۱۰] (توری[۱۱]) با اندازهٔ $\boldsymbol{n} = (n_1, n_2)$ است. مجموعهٔ $\mathcal{I} = \{1, \ldots, n\}$ نیز نشانگر موقعیت‌های فضایی است، به طوری که موقعیت‌های با مرز مشترک همسایه‌اند و مشبکهٔ نامنظم[۱۲] نامیده می‌شود. برای هر $C \subset \mathcal{I}$، $\boldsymbol{x}_C$ به‌صورت $\boldsymbol{x}_C = \{x_i : i \in C\}$ تعریف می‌شود. اگر مجموعهٔ $\mathcal{I} - C$ با $-C$ نشان داده شود، آن‌گاه $\boldsymbol{x}_{-C} = \{x_i : i \in -C\}$. برای دو زیرمجموعهٔ C_1 و C_2 از $\mathcal{I}$،

[۹] Leading Principal Submatrix
[۱۰] Regular Lattice
[۱۱] Grid
[۱۲] Irregular Lattice

$C_۱ \backslash C_۲ = \{i \in \mathcal{I} : i \in C_۱, i \notin C_۲\}$ است.

تعداد عملگرهای محاسبه یک عبارت، فلاپ (flop)[۱۳] نامیده می‌شود. به عنوان مثال، عبارت $x + a * b$ که شامل یک عملگر ضرب و یک عملگر جمع است، دارای دو فلاپ است.

۱.۲.۱ همبستگی فضایی

وجود همبستگی فضایی داده‌ها با استفاده از آزمون I—موران[۱۴] (موران، ۱۹۵۰) قابل بررسی است. در این آزمون مقادیر متغیر در هر موقعیت با سایر موقعیت‌ها از طریق شاخص

$$I = \frac{n \sum_{i=۱}^{n} \sum_{j=۱}^{n} w_{ij}(y_i - \overline{y})(y_j - \overline{y})}{(\sum_{i=۱}^{n} \sum_{j=۱}^{n} w_{ij}) \sum_{i=۱}^{n} (y_i - \overline{y})^۲}$$

مقایسه می‌شوند، که در آن y_i و y_j به ترتیب مقادیر متغیر پاسخ در موقعیت‌های i و j، w_{ij} درایه‌های ماتریس وزن همسایگی $\boldsymbol{W}$ هستند، که به یکی از روش‌های زیر انتخاب می‌شوند:

الف— w_{ij} معکوس فاصلهٔ اقلیدسی بین دو موقعیت i و j است.

ب— اگر بین دو موقعیت i و j مرز مشترک وجود داشته باشد، w_{ij} برابر یک و در غیر این‌صورت صفر است.

فرضیه‌های آزمون موران به‌صورت

$$\begin{cases} H_۰ : \text{عدم وجود همبستگی فضایی} \\ H_۱ : \text{وجود همبستگی فضایی} \end{cases}$$

و آمارهٔ آزمون و توزیع آن به‌صورت

$$Z(I) = \frac{I - E(I)}{S_{E(I)}} \sim \mathcal{N}(۰, ۱)$$

---

[۱۳] Floating Point Operation

[۱۴] Moran-I Test

است، که در آن

$$S^{۲}_{E(I)} = \frac{n^{۲}\sum_i\sum_j w^{۲}_{ij} + ۳(\sum_i\sum_j w_{ij})^{۲} - n\sum_i(\sum_j w_{ij})^{۲}}{(n^{۲}-۱)(\sum_i\sum_j w_{ij})^{۲}}.$$

تعریف ۱.۲.۱ (استقلال شرطی): دو متغیر تصادفی x و y برای z داده شده مستقل شرطی نامیده می‌شوند، اگر و تنها اگر $\pi(x,y|z) = \pi(x|z)\pi(y|z)$ و با نماد $x \perp y|z$ نمایش داده می‌شود.

۲.۲.۱ گراف

مجموعه‌ای از رأس‌ها که توسط تعدادی یال به هم وصل شده‌اند، گراف نامیده می‌شود. گراف[۱۵] $\mathcal{G}$ به‌صورت زوج مرتب $(\mathcal{V},\mathcal{E})$ نشان داده می‌شود، که در آن $\mathcal{V}$ مجموعه‌ای متناهی و غیر تهی از رأس‌ها و $\mathcal{E}$ زیر مجموعهٔ تمام زوج‌های نامرتب $\{i,j\}$ از عناصر متمایز $\mathcal{V}$ به‌صورت

$$\mathcal{E} = \{\{i,j\} : i,j \in \mathcal{V}, i \neq j\}$$

است. اعضای $\mathcal{V}$ را رأس‌های گراف $\mathcal{G}$ و اعضای $\mathcal{E}$ را یال‌های آن می‌نامند. بین دو رأس یک گراف ساده[۱٦] حداکثر یک یال وجود دارد. شکل ۱.۲.۱ یک گراف ساده شامل چهار رأس و سه یال را نشان می‌دهد.

گراف $\mathcal{G}$ همبند کامل[۱۷] نامیده می‌شود هرگاه بین هر دو رأس آن دقیقاً یک یال وجود داشته باشد. اگر $\mathcal{E}$ زیر مجموعه‌ای از تمام زوج‌های مرتب عناصر $\mathcal{V}$ باشد، یعنی

$$\mathcal{E} \subseteq \{(i,j) : i,j \in \mathcal{V}, i \neq j\}$$

---

[۱۵] Graph

[۱٦] Simple Graph

[۱۷] Full Connected