

الله أكبر

وزارت علوم، تحقیقات و فناوری

دانشگاه بین‌المللی امام خمینی



دانشکده علوم

تحلیل رگرسیونی داده‌های شمارشی با متغیر پاسخ آمیخته

پایان‌نامه برای دریافت درجه کارشناسی ارشد
در رشته‌ی آمار گرایش آمار ریاضی

مهسا نادى فر

استاد راهنما:

دکتر افشین فلاح

دی ماه ۱۳۹۲

وزارت علوم، تحقیقات و فناوری

دانشگاه بین‌المللی امام خمینی



دانشکده علوم
گروه آمار

تحلیل رگرسیونی داده‌های شمارشی با متغیر پاسخ آمیخته

پایان‌نامه برای دریافت درجه کارشناسی ارشد در رشته‌ی آمار
گرایش آمار ریاضی

مهسا نادى‌فر

استاد راهنما:

دکتر افشین فلاح

استاد مشاور:

دکتر رامین کاظمی

دی ماه ۱۳۹۲

من اینک در ابتدای راه آموختن فراتر از آن چه آموختم،

سوگند می خورم به نام مقدس عشق،

سوگند به محطه های خاموش صبرمادر،

سوگند به گرمی دستان استوار پدر،

آن چه امروز در برابرش ایستاده ام کمترین اندوخته ایست که با هر چه نامش را عشق و

صبر و استواری می نامند، به تو پدر و به تو مادر تقدیم می کنم.

و تقدیم به کسانی که از سرمایه وجودی علم خود بی دریغ بخشیدند
تا سهمی از نور راد خود جای نهند...

« نفس های تان پناه راه من است. »

سپاس گزارى

دستى كه كشاده است
مى برد،
مى آورد،
رهنمونى مى شود
به خانه نى كه
نور دى پيش كرمى بخش است.

سپاس خدايى را كه اول است و پيش از او اولى نبوده و آخر است و آخرى نباشد.
منت خداى عزوجل بر من است كه توفيق راه يابى به وادى علم را عطا فرمود. اکنون كه محقق شده است رسيدن به
پايان سرفصل، بر خود واجب مى دانم كه ارج نهم بزرگانى را كه آورده هاى شمع وجودشان روشنگر راهم است.
سپاس بى شائبه ي خود را تقديم مى دارم به محضر استاد ارجمندم جناب آقاى دكتر افشين فلاح كه همواره رهنمودهايشان
يارى دهنده ي مشكلات علمى ام بوده است. تقدير و تشكر خود را از استاد مشاور جناب آقاى دكتر اسين كاظمى صميانه
ابراز مى دارم. از استاد ارجمند، جناب آقاى دكتر امير كاوسى جهت تهيه داده هاى مربوط به سرطان معده و روده،
سپاس گزارم. از استاد كراتقدر، جناب آقاى دكتر اسماعيل اميرى جهت داوري اين پايان نامه كمال تشكر را دارم.
در پايان از پدر و مادر مهربانم و برادران عزيزم يشم و مرتضى كمال تشكر را دارم.

مسانادى فر

فرم پ ۲-۱: فرم تأییدیه هیأت داوران جلسه دفاع از پایان نامه

دانشگاه بین المللی امام خمینی



IMAM KHOMEINI INTERNATIONAL UNIVERSITY

دانشگاه بین المللی امام خمینی (ره)

معاونت آموزشی - مدیریت تحصیلات تکمیلی

فرم تأییدیه هیأت داوران جلسه دفاع از پایان نامه / رساله (فرم شماره ۳۰)

بدین وسیله گواهی میشود جلسه دفاعیه از پایان نامه کارشناسی ارشد مهسا نادی فر دانشجوی رشته آمار گرایش آمار ریاضی تحت عنوان تحلیل رگرسیونی داده‌های شمارشی با متغیر پاسخ آمیخته در تاریخ ۰۷ / ۱۰ / ۱۳۹۲ در دانشگاه برگزار گردید و این پایان نامه با نمره به عدد ۲۰ و به حروف بیست با درجه عالی مورد تأیید هیئت داوران قرار گرفت.

ردیف	سمت	نام و نام خانوادگی	مرتبه‌ی دانشگاهی	دانشگاه یا مؤسسه	امضا
۱	استاد راهنما	آقای دکتر افشین فلاح	استادیار	دانشگاه بین المللی امام خمینی (ره)	
۲	استاد مشاور	آقای دکتر رامین کاظمی	استادیار	دانشگاه بین المللی امام خمینی (ره)	
۳	داور	آقای دکتر اسماعیل امیری	استادیار	دانشگاه بین المللی امام خمینی (ره)	
۴	نماینده تحصیلات تکمیلی	آقای دکتر ابراهیم وطن دوست	استادیار	دانشگاه بین المللی امام خمینی (ره)	



تعهدنامه اصالت اثر

اینجانب مهسا نادى فر دانش آموخته مقطع کارشناسی ارشد در رشته آمار گرایش آمار ریاضی که در تاریخ ۹۲/۱۰/۰۷ از پایان نامه ی خود تحت عنوان تحلیل رگرسیونى داده‌های شمارشی با متغیر پاسخ آمیخته با کسب درجه ی عالی دفاع کرده ام، شرعا و قانونا متعهد می شوم:

۱. مطالب مندرج در این پایان نامه، حاصل تحقیق و مطالعه اینجانب بوده و در مواردی که از دستاوردهای علمی و پژوهشی دیگران اعم از پایان نامه، کتاب، مقاله و غیره استفاده کرده ام، با رعایت کامل امانت، مطابق مقررات، اقدام به ارجاع در متن و ذکر آن در فهرست منابع و مآخذ نموده ام.
۲. تمامی یا بخشی از این پایان نامه قبلا برای دریافت هیچ مدرک تحصیلی به سایر دانشگاه ها و موسسات آموزش عالی ارائه نشده است.
۳. مقالات مستخرج از این پایان نامه کاملا حاصل کار اینجانب بوده و از هرگونه جعل داده و یا تغییر اطلاعات پرهیز کرده ام.
۴. از ارسال همزمان و یا تکراری مقالات مستخرج از این پایان نامه (با بیش از ۳ درصد همپوشانی) به مجلات و یا همایش های گوناگون خودداری نموده و می نمایم.
۵. کلیه حقوق مادی و معنوی حاصل از این پایان نامه متعلق به دانشگاه بین المللی امام خمینی (ره) بوده و متعهد می شوم هرگونه بهره مندی ویا نشر دستاوردهای حاصل از این تحقیق اعم از چاپ کتاب، مقاله، ثبت اختراع و غیره (چه در زمان دانشجویی و یا بعد از فراغت از تحصیل) با کسب اجازه از استاد (استادان) رهنما باشد.
۶. در صورت اثبات تخلف و نقض موارد پنجگانه فوق (در هر زمان) مدرک تحصیلی صادر شده توسط دانشگاه بین المللی امام خمینی (ره) از درجه اعتبار ساقط و اینجانب هیچگونه ادعایی نخواهم داشت.

نام و نام خانوادگی دانشجو

مهسا نادى فر

امضاء



سوگندنامه دانش آموختگان کارشناسی ارشد دانشگاه بین المللی امام خمینی (ره)

به نام خدا

سپاس ایزد منان را که مرا مشمول الطاف خویش نمود که با طی مراحل تحصیل موفق به اخذ درجه کارشناسی ارشد شوم. به شکرانه این نعمت بزرگ الهی که با امکانات این مرز و بوم، فراهم و نزد اینجانب به امانت گذاشته شده است، در پیشگاه ملت ایران به کتاب آسمانی خود، قرآن کریم، سوگند یاد می کنم که:

- در سراسر زندگی حرفه ای، در راه اعتلای کشور ایران و جامعه بشری به نحو احسن قدم برداشته و در این راه از هیچ تلاشی دریغ ننمایم.
- در تمام فعالیت های تخصصی، رضای خدا را همراه با صداقت علمی و اجتماعی در نظر داشته و از موقعیت های به دست آمده در جهت رفع مشکلات جامعه استفاده کنم و در همه ی امور، منافع کشور را بر منافع فردی مقدم بدارم.
- همواره علم و دانش خود را به روز نگاه داشته و در ایفای مسئولیت و تعهدات حرفه ای در حد توان سعی و تلاش خود را به کار گیرم.
- و اینک از خداوند علیم توفیق بندگی و پای بندی به مفاد این سوگندنامه را خواستارم و از او می خواهم که مرا در ایفای رسالت علمی و انسانی خویش موفق بدارد.

نام و نام خانوادگی دانشجو

مسائری

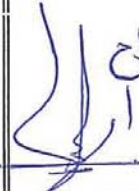
امضاء

مجوز بهره برداری از پایان نامه/رساله

کلیه حقوق اعم از چاپ، تکثیر، نسخه برداری، ترجمه، اقتباس و ... از نتایج این پایان نامه برای دانشگاه بین المللی امام خمینی (ره) قزوین محفوظ است. بهره برداری از این پایان نامه/رساله در چهارچوب مقررات کتابخانه و با توجه به محدودیتی که توسط استاد راهنما به شرح زیر تعیین می شود، بلامانع است:

- بهره برداری از این پایان نامه/رساله برای همگان بلامانع است.
- بهره برداری از این پایان نامه/رساله با اخذ مجوز از استاد راهنما، بلامانع است.
- بهره برداری از این پایان نامه/رساله تا تاریخ ممنوع است.

استاد راهنما می تواند یکی از گزینه های بالا را انتخاب کند و مسئولین کتابخانه موظف به رعایت موارد تعیین شده می باشد.

نام استاد و یا اساتید راهنما: افسن بلاح
تاریخ: ۹۲، ۱۲، ۲۴
امضاء: 

چکیده

در این پایان‌نامه تحلیل رگرسیونی داده‌های شمارشی با متغیر پاسخ آمیخته مدنظر قرار گرفته است. در تحلیل داده‌های شمارشی، تحلیل رگرسیون پواسون یکی از تکنیک‌های بسیار پرکاربرد می‌باشد. یکی از مفروضات اصلی مدل رگرسیون پواسون برابری میانگین و واریانس متغیر پاسخ است. اما در بسیاری از کاربردها مشاهدات پاسخ بیش‌پراکنده‌اند و به همین دلیل برازش مدل رگرسیون پواسون به داده‌ها مناسب نیست. در این موارد می‌توان از مدل رگرسیون پواسون آمیخته به‌عنوان مدلی جایگزین استفاده نمود. از این رو، مدل رگرسیون پواسون آمیخته‌ی متناهی یک متغیره از دو دیدگاه بسامدی و بیزی مورد بررسی قرار گرفته است. سپس کاربرد مدل رگرسیون پواسون آمیخته‌ی متناهی در تحلیل رگرسیونی داده‌های پیوندیافته شرح داده شده است. وجود صفرهای بیش از حد در مشاهدات پاسخ یکی دیگر از مشکلات در تحلیل داده‌های شمارشی است، که می‌توان با استفاده از توزیع پواسون دومتغیره بر آن غلبه کرد. بنا بر این تحلیل رگرسیونی با متغیر پاسخ دارای توزیع پواسون آمیخته دو متغیره در رهیافت بیزی مدنظر قرار گرفته است. نشان داده شده که به دلیل شکل پیچیده تابع درستنمایی مبتنی بر توزیع پواسون دو متغیره، توزیع پسین فاقد شکل بسته بوده و بسیار پیچیده است. از این رو، توزیع‌های پسین شرطی کامل پارامترها محاسبه و الگوریتم گیبز برای نمونه‌گیری از توزیع پسین ارائه شده است. به‌منظور ارزیابی مدل بیزی پیشنهادی و مقایسه‌ی آن با همتای بسامدی، مطالعه‌ای شبیه‌سازی انجام شده است. هم‌چنین نحوه کاربست رهیافت بیزی پیشنهادی در قالب یک مثال کاربردی در زمینه‌ی تحلیل داده‌های سرطان معده و روده‌ی بزرگ شرح داده شده و کارایی آن مورد ارزیابی قرار گرفته است. نتایج حاکی از کارایی مطلوب‌تر مدل بیزی پیشنهادی نسبت به مدل بسامدی در برازش مدل است.

واژه‌های کلیدی: رگرسیون پواسون، توزیع آمیخته، رهیافت بیزی، الگوریتم EM ، الگوریتم متروپولیس-هاستینگز، الگوریتم گیبز.

فهرست مطالب

ر	فهرست مطالب
ش	فهرست جدول‌ها
پ	فهرست شکل‌ها
۱	۱ مقدمه‌ای بر تحلیل بسامدی و بیزی توزیع‌های آمیخته
۱	۱.۱ مقدمه
۱	۲.۱ مقدمه‌ای بر مدل‌های رگرسیونی شمارشی
۵	۳.۱ توزیع‌های آمیخته
۷	۱.۳.۱ شناسایی‌پذیری توزیع‌های آمیخته
۱۰	۲.۳.۱ تابع درست‌نمایی داده‌های ناقص
۱۱	۳.۳.۱ تابع درست‌نمایی داده‌های کامل
۱۳	۴.۳.۱ گشتاورهای یک توزیع آمیخته
۱۴	۵.۳.۱ برآورد پارامترهای توزیع آمیخته در رهیافت بسامدی با استفاده از الگوریتم EM
۱۷	۶.۳.۱ تحلیل بیزی توزیع‌های آمیخته
۲۱	۲ تحلیل رگرسیونی داده‌های شمارشی با متغیر پاسخ آمیخته
۲۱	۱.۲ مقدمه

۲۱	توزیع پواسون	۲.۲
۲۴	برآورد پارامتر توزیع پواسون	۱.۲.۲
۲۴	رابطه‌ی توزیع پواسون با سایر توزیع‌ها	۲.۲.۲
۲۷	توزیع پواسون آمیخته	۳.۲
۲۷	برآورد پارامترهای توزیع پواسون آمیخته با استفاده از الگوریتم <i>EM</i>	۱.۳.۲
۲۸	مدل‌های خطی تعمیم‌یافته	۴.۲
۲۹	مدل رگرسیون پواسون	۵.۲
۳۰	برآورد ماکسیمم درست‌نمایی پارامترهای مدل رگرسیون پواسون	۱.۵.۲
۳۱	برآورد ماکسیمم درست‌نمایی پارامترهای مدل رگرسیون پواسون با استفاده از الگوریتم کیم-زیم	۲.۵.۲
۳۴	بررسی مناسبت مدل	۳.۵.۲
۳۵	ضعف‌های مدل رگرسیون پواسون	۴.۵.۲
۳۶	مدل رگرسیون دوجمله‌ای منفی	۶.۲
۳۷	مدل رگرسیون آمیخته‌ی متناهی پواسون	۷.۲
۳۹	تحلیل رگرسیونی پواسون داده‌های پیوند یافته	۱.۷.۲
۴۵	مطالعه‌ی شبیه‌سازی	۸.۲
۵۱	تحلیل رگرسیونی پواسون آمیخته‌ی متناهی دومتغیره	۳
۵۱	مقدمه	۱.۳
۵۱	توزیع پواسون دومتغیره	۲.۳
۵۳	مدل رگرسیون پواسون دومتغیره	۱.۲.۳
۵۴	توزیع پواسون آمیخته‌ی دومتغیره	۲.۲.۳
۵۶	توزیع پواسون آمیخته‌ی متناهی دومتغیره	۳.۲.۳
۶۰	مدل رگرسیون پواسون آمیخته‌ی متناهی دومتغیره	۴.۲.۳

۶۱	۳.۳ رهیافت بسامدی برازش مدل رگرسیون پواسون آمیخته متناهی دومتغیره	
۶۵	۴.۳ رهیافت بیزی برازش مدل رگرسیون پواسون آمیخته‌ی متناهی دومتغیره	
۶۶	۱.۴.۳ توزیع‌های پیشین	
۶۷	۲.۴.۳ توزیع پسین	
۶۸	۳.۴.۳ توزیع‌های پسین شرطی کامل	
۷۱	۵.۳ مطالعه‌ی شبیه‌سازی	
۷۳	۶.۳ تحلیل داده‌های سرطان روده و معده	
۸۱	۴ توزیع پواسون آمیخته‌ی متناهی چندمتغیره	
۸۱	۱.۴ مقدمه	
۸۱	۲.۴ مقدمه‌ای بر توزیع پواسون آمیخته‌ی متناهی چندمتغیره	
۸۳	۳.۴ توزیع پواسون چندمتغیره	
۸۵	۴.۴ توزیع پواسون آمیخته‌ی چندمتغیره	
۸۶	۱.۴.۴ توزیع پواسون آمیخته‌ی متناهی چندمتغیره	
۸۷	۲.۴.۴ توزیع‌های حاشیه‌ای و شرطی توزیع پواسون آمیخته‌ی متناهی چندمتغیره	
۸۸	۳.۴.۴ استنباط در مورد توزیع پواسون آمیخته‌ی متناهی چندمتغیره	
۸۹	۵.۴ بحث و نتیجه‌گیری	
۹۱		پیوست
۹۹		پیوست
۱۰۵		پیوست
۱۱۵		پیوست

۱۲۲

پیوست

۱۲۵

واژه‌نامه

۱۲۸

نام‌نامه

۱۳۲

مراجع

فهرست جدول‌ها

- ۱.۲ مقادیر کم‌ترین توان‌های دوم خطا برای براوردگرهای بیزی و کم‌ترین توان‌های دوم وزنی تکراری. ۴۸
- ۲.۲ مقدار آماره ارزیابی متقابل برای پیش‌بینی‌های حاصل از براوردگرهای بیزی و کم‌ترین توان‌های دوم وزنی تکراری. ۴۹
- ۱.۳ کارایی نسبی براوردگر بسامدی نسبت به براوردگر بیز پیشنهادی در براورد ضرایب رگرسیونی به‌ازای مقادیر مختلف ضرایب آمیختگی در مولفه اول. ۷۴
- ۲.۳ کارایی نسبی براوردگر بسامدی نسبت به براوردگر بیز پیشنهادی در براورد ضرایب رگرسیونی به‌ازای مقادیر مختلف ضرایب آمیختگی در مولفه دوم. ۷۵
- ۳.۳ کارایی نسبی براوردگر بسامدی به براوردگر بیز پیشنهادی در براورد ضرایب آمیختگی به‌ازای مقادیر مختلف این پارامتر. ۷۷
- ۴.۳ تعداد کل موارد بروز سرطان‌های معده و روده بزرگ در سال‌های ۱۳۸۳ تا ۱۳۸۷ و عوامل مؤثر بر این سرطان‌ها در سال ۱۳۸۷. ۷۸
- ۵.۳ خلاصه آمارهای توصیفی متغیرهای پاسخ و تبیینی. ۷۸
- ۶.۳ براورد پارامترهای مدل رگرسیون پواسون دومتغیره آمیخته در دو رهیافت بیزی و بسامدی. ۷۹
- ۷.۳ مقدار آماره ارزیابی متقابل برای پیش‌بینی‌های حاصل از دو رهیافت بیزی و بسامدی. ۷۹

فهرست شکل‌ها

- ۱.۲ نمودارهای اثر نمونه‌های حاصل از توزیع‌های پسین ضرایب رگرسیونی β_1 (ستون اول) و β_2 (ستون دوم)، به‌ازای خطاهای A_1 (ردیف اول)، A_2 (ردیف دوم) و A_3 (ردیف سوم). ۴۷
- ۲.۲ نمودارهای خودهمبستگی نمونه‌های حاصل از توزیع‌های پسین ضرایب رگرسیونی β_1 (ستون اول) و β_2 (ستون دوم)، به‌ازای خطاهای A_1 (ردیف اول)، A_2 (ردیف دوم) و A_3 (ردیف سوم). ۴۷
- ۱.۳ نمودارهای اثر (ب) و خودهمبستگی (الف) نمونه‌های حاصل از توزیع‌های پسین ضرایب رگرسیونی β_{111} (ردیف اول)، β_{211} (ردیف دوم)، β_{311} (ردیف سوم)، β_{121} (ردیف چهارم)، β_{221} (ردیف پنجم) و β_{321} (ردیف ششم). ۷۶

فصل ۱

مقدمه‌ای بر تحلیل بسامدی و بیزی توزیع‌های آمیخته

۱.۱ مقدمه

در این فصل به بیان مقدمه‌ای درباره‌ی تحلیل رگرسیون داده‌های شمارشی و مدل‌های رگرسیونی با متغیر پاسخ آمیخته پرداخته شده است. همچنین مفاهیم اساسی مرتبط با توزیع‌های آمیخته و برآورد پارامترهای این توزیع‌ها در دو رهیافت بسامدی و بیزی شرح داده شده است.

۲.۱ مقدمه‌ای بر مدل‌های رگرسیونی شمارشی

تحلیل رگرسیون پواسون از تکنیک‌های بسیار پرکاربرد در زمینه‌ی تحلیل داده‌های شمارشی است. مدل رگرسیون پواسون نوعی مدل خطی تعمیم‌یافته است که در آن متغیر وابسته، شمارشی بوده و از توزیع پواسون پیروی می‌کند. در تحلیل رگرسیون پواسون، میانگین یک متغیر شمارشی با نرخ وقوع خاص از طریق تابع پیوند نمایی به مجموعه‌ای از متغیرهای تبیینی ارتباط داده می‌شود. برابری واریانس متغیر پاسخ با میانگین آن یکی از مفروضات اصلی تحلیل رگرسیون پواسون است. اما در بسیاری از موارد مشاهدات پاسخ بیش‌پراکنده‌اند، به این معنی که واریانس مشاهدات به طور معنی‌داری از میانگین آن‌ها بزرگ‌تر می‌باشد. همچنین ممکن است پارامتر نرخ وقوع در توزیع پواسون خود یک متغیر تصادفی باشد. در این شرایط برازش مدل رگرسیون پواسون به داده‌ها مناسب نیست. در این مواقع می‌توان از تحلیل رگرسیون دو جمله‌ای

منفی کمک گرفت (گاردنر و همکاران، ۱۹۹۵). توزیع دوجمله‌ای منفی، گسسته و نامنفی بوده و بر خلاف توزیع پواسون نسبت به برابری میانگین و واریانس مشاهدات حساس نیست. از این رو، به دلیل تشابه تکیه‌گاه این دو توزیع، می‌توان از آن به‌عنوان جایگزینی مناسب برای توزیع متغیر پاسخ در رگرسیون پواسون استفاده کرد.

یکی از راه‌های جایگزین در شرایط بیش‌پراکنش مشاهدات پاسخ در تحلیل رگرسیونی داده‌های شمارشی استفاده از توزیع‌های آمیخته است. توزیع‌های آمیخته در زمینه‌های مختلفی از جمله ستاره‌شناسی، زیست‌شناسی، ژنتیک، پزشکی، روان‌پزشکی، اقتصاد، مهندسی، تجارت و غیره به‌صورت گسترده مورد استفاده قرار می‌گیرند. در سال‌های اخیر برآورد پارامترهای توزیع آمیخته از دو دیدگاه بسامدی و بیزی و هم‌چنین آزمون مرتبط با تعداد مؤلفه‌های یک توزیع آمیخته، توجه بسیاری از محققان را به خود معطوف ساخته است.

توزیع‌های آمیخته اولین بار توسط پیرسون (۱۸۹۴) در قالب یک توزیع نرمال آمیخته‌ی متناهی دو مؤلفه‌ای مطرح شدند. در دهه‌ی ۱۹۶۰ محققینی از جمله وولف (۱۹۶۵) برآورد ماکسیمم درست‌نمایی توزیع‌های آمیخته را مورد مطالعه‌ی اولیه قرار دادند. سپس دمپستر و همکاران (۱۹۷۷) الگوریتم EM را برای برآورد ماکسیمم درست‌نمایی مدل‌های آمیخته ارائه کردند. یکی از مسائل اصلی در توزیع‌های آمیخته تعیین تعداد مؤلفه‌های توزیع است. آکائیکه (۱۹۷۳) به کمک نظریه‌ی اطلاع، معیار اطلاع آکائیکه را برای تعیین تعداد مؤلفه‌های یک توزیع آمیخته، از طریق مینیمم ساختن فاصله‌ی کولبک-لیبلر بین توزیع جامعه و توزیع‌های پیشنهادی، مورد مطالعه قرار داد. رهیافت بیزی تعیین تعداد مؤلفه‌های توزیع آمیخته، توسط ریچاردسون و گرین (۱۹۹۷) مورد بررسی قرار گرفت. توزیع‌های رگرسیون آمیخته‌ی داده‌های شمارشی به دلیل انعطاف‌پذیر بودن در مدل‌بندی بسیاری از پدیده‌های تصادفی مفید هستند و در بسیاری از تحقیقات کاربردی به صورت گسترده مورد استفاده قرار می‌گیرند. مدل رگرسیون آمیخته‌ی پواسون را می‌توان بر داده‌های شمارشی بیش‌پراکنده برازش داد. وانگ و همکاران (۱۹۹۸) مدل رگرسیون پواسون آمیخته با متغیرهای تبیینی متفاوت در هر مؤلفه‌ی آمیخته را در نظر گرفتند. بوهنینگ (۱۹۹۹) نیز تحلیل و کاربرد توزیع‌های آمیخته را در حالت کلی مورد مطالعه قرار داد. از جمله کتاب‌های نوشته شده در زمینه‌ی کاربرد توزیع‌های آمیخته می‌توان به لیندسی (۱۹۹۵) و مک‌لاچلان و پیل (۲۰۰۰) اشاره کرد، که به تحلیل توزیع‌های آمیخته‌ی متناهی به‌عنوان توزیع‌های مناسب در موقعیت‌های گوناگون و برآورد پارامترهای این توزیع‌ها در دو رهیافت بسامدی و بیزی و هم‌چنین تعیین تعداد

مؤلفه‌های یک توزیع آمیخته پرداخته‌اند. فرنوش و همکاران (۲۰۰۹) آزمون همگنی توزیع‌های آمیخته را ارائه کردند. وجود صفرهای بیش از حد در مشاهدات متغیر پاسخ نیز یکی دیگر از مشکلات در تحلیل داده‌های شمارشی است. لمبرت (۱۹۹۲) مدل رگرسیون پواسون در صفر متورم را معرفی کرد. مدل رگرسیون دوجمله‌ای منفی در صفر متورم نیز توسط وانگ (۲۰۰۳) و گارای و همکاران (۲۰۱۱) مورد مطالعه قرار گرفته است. یکی از راه‌های مقابله با مشکل وجود صفرهای بیش از حد در مشاهدات پاسخ، استفاده از توزیع پواسون دومتغیره است. مدل‌های در صفر متورم در حالت دو متغیره توسط کارلیس و ان‌سوفراس (۲۰۰۳) و گورمو و الدر (۲۰۰۸) مورد مطالعه قرار گرفته‌اند. اخیراً برمودز (۲۰۰۹) و برمودز و کارلیس (۲۰۱۱) درباره‌ی نحوه استفاده از مدل‌های رگرسیون در صفر متورم دو و چندمتغیره برای رتبه‌بندی بیمه اتومبیل مطالعاتی انجام داده‌اند. آن‌ها مدل رگرسیون پواسون دومتغیره را برای بررسی همبستگی ذاتی دو ادعا در صنعت بیمه مورد استفاده قرار دادند. یکی از مزیت‌های استفاده از توزیع‌های دومتغیره این است که در این توزیع‌ها می‌توان همبستگی بین دو متغیر پاسخ مختلف را در نظر گرفت و به نتایج مطلوب‌تری دست یافت. گرچه این مدل‌ها برای داده‌های در صفر متورم مناسب هستند، اما برای مدل‌بندی مشاهداتی که بیش‌پراکنده نیز باشند، از کارایی مطلوبی برخوردار نیستند. از این رو، برمودز و کارلیس (۲۰۱۲) مدل رگرسیون آمیخته‌ی متناهی پواسون دومتغیره را مطرح و نحوه‌ی برآورد پارامترهای مدل را بر اساس رهیافت بسامدی ماکسیمم درست‌نمایی و به کمک الگوریتم *EM* شرح دادند. این مدل نسبت به مدل‌های دیگر دارای انعطاف‌پذیری بیشتری بوده و قابلیت مدل‌بندی بیش‌پراکنش و همبستگی منفی را نیز دارا است. این مدل در زمینه‌های مختلفی مانند رتبه‌بندی بیمه‌ها، جرم‌شناسی و ژنتیک از اهمیت بالایی برخوردار بوده و بسیار پرکاربرد می‌باشد. هر چند تحلیل داده‌های پیوسته‌ی چندمتغیره دارای سابقه‌ای طولانی است، اما در مورد داده‌های شمارشی چندمتغیره، به دلیل پیچیده بودن ساختار توزیع‌های شمارشی چندمتغیره تا سال‌های اخیر مطالعات چندانی صورت نگرفته است. گرچه می‌توان توزیع‌های شمارشی چندمتغیره را توسط توزیع نرمال چندمتغیره تقریب زد، اما این تقریب به از دست رفتن اطلاعات و تضعیف تحلیل حاصل از آن منجر می‌شود. جانسون و همکاران (۱۹۹۷) توزیع پواسون چندمتغیره را که بین توزیع‌های شمارشی چندمتغیره از اهمیت ویژه‌ای برخوردار است، معرفی کردند. مشکل اصلی در استفاده از توزیع پواسون چندمتغیره پیچیده بودن شکل توزیع آن است. برای مطالعه‌ی بیش‌تر در خصوص فرایندهای استنباطی راجع به توزیع پواسون چندمتغیره می‌توان به تیوناس (۱۹۹۹) و کارلیس (۲۰۰۳) مراجعه

کرد. مدل رگرسیون پواسون چندمتغیره با ساختار کواریانس مشخص بین متغیرهای پاسخ توسط کارلیس و ملیگکوتسیدو (۲۰۰۳) ارائه شده است. این مدل به دلیل در نظر گرفتن متغیرهای تبیینی در مقایسه با مدل رگرسیون پواسون چندمتغیره از انعطاف پذیری بیشتری برخوردار است. توزیع پواسون چندمتغیره نیز در حالتی که داده‌های تحت مطالعه بیش پراکنده باشند، مناسب نیست. هم‌چنین این توزیع قابلیت مدل‌بندی همبستگی منفی بین متغیرها را دارا نیست، از این رو، می‌توان از توزیع آمیخته‌ی متناهی پواسون چندمتغیره استفاده کرد. کارلیس و ملیگکوتسیدو (۲۰۰۷) توزیع آمیخته‌ی متناهی پواسون چندمتغیره را معرفی کردند و برآورد پارامترهای این توزیع را در رهیافت بسامدی با استفاده از الگوریتم EM انجام داده و کاربرد توزیع را در زمینه‌ی جرم‌شناسی نشان دادند. برای سهولت در امر محاسبات می‌توان از توزیع‌های شمارشی دومتغیره به عنوان حالت خاصی از توزیع‌های چندمتغیره استفاده کرد.

در این پایان‌نامه توزیع‌های شمارشی پواسون و دوجمله‌ای منفی و برخی خصوصیات این دو توزیع مورد بحث قرار گرفته است. هم‌چنین مدل‌های رگرسیونی شمارشی آمیخته از جمله مدل رگرسیون پواسون یک‌متغیره با متغیر پاسخ آمیخته‌ی متناهی و ویژگی‌های آن بیان شده است. چگونگی برآورد پارامترهای مدل رگرسیون پواسون آمیخته‌ی متناهی در دو رهیافت بسامدی و بیزی به ترتیب با استفاده از الگوریتم‌های EM و زنجیرهای مارکف مونت کارلویی، شرح داده شده است. هم‌چنین کارایی این مدل برای داده‌های پیوندیافته در قالب یک مطالعه‌ی شبیه‌سازی ارزیابی شده است. در فصل سوم، ابتدا چگونگی ساختن توزیع پواسون دومتغیره و توزیع پواسون آمیخته‌ی متناهی دومتغیره بیان و برخی ویژگی‌های آن‌ها شرح داده شده است. سپس، مدل رگرسیون پواسون با متغیر پاسخ آمیخته‌ی متناهی دومتغیره و نحوه‌ی برآورد پارامترهای این مدل در دو رهیافت بسامدی و بیزی مورد مطالعه قرار گرفته است. هم‌چنین به منظور ارزیابی مدل پیشنهادی در رهیافت بیزی، مطالعه‌ای شبیه‌سازی انجام شده است. به علاوه نحوه‌ی کاربست این مدل در دو رهیافت بسامدی و بیزی در قالب یک مثال کاربردی در زمینه مطالعات پزشکی شرح داده شده است. در فصل چهارم، نحوه‌ی ساختن توزیع پواسون چندمتغیره و توزیع پواسون آمیخته‌ی متناهی چندمتغیره بیان شده است. سپس پارامترهای یک توزیع پواسون آمیخته‌ی متناهی چندمتغیره در رهیافت بسامدی با استفاده از الگوریتم EM برآورد شده است. در پایان این فصل نیز نتیجه‌گیری کلی از تمامی فصول ارائه شده است.