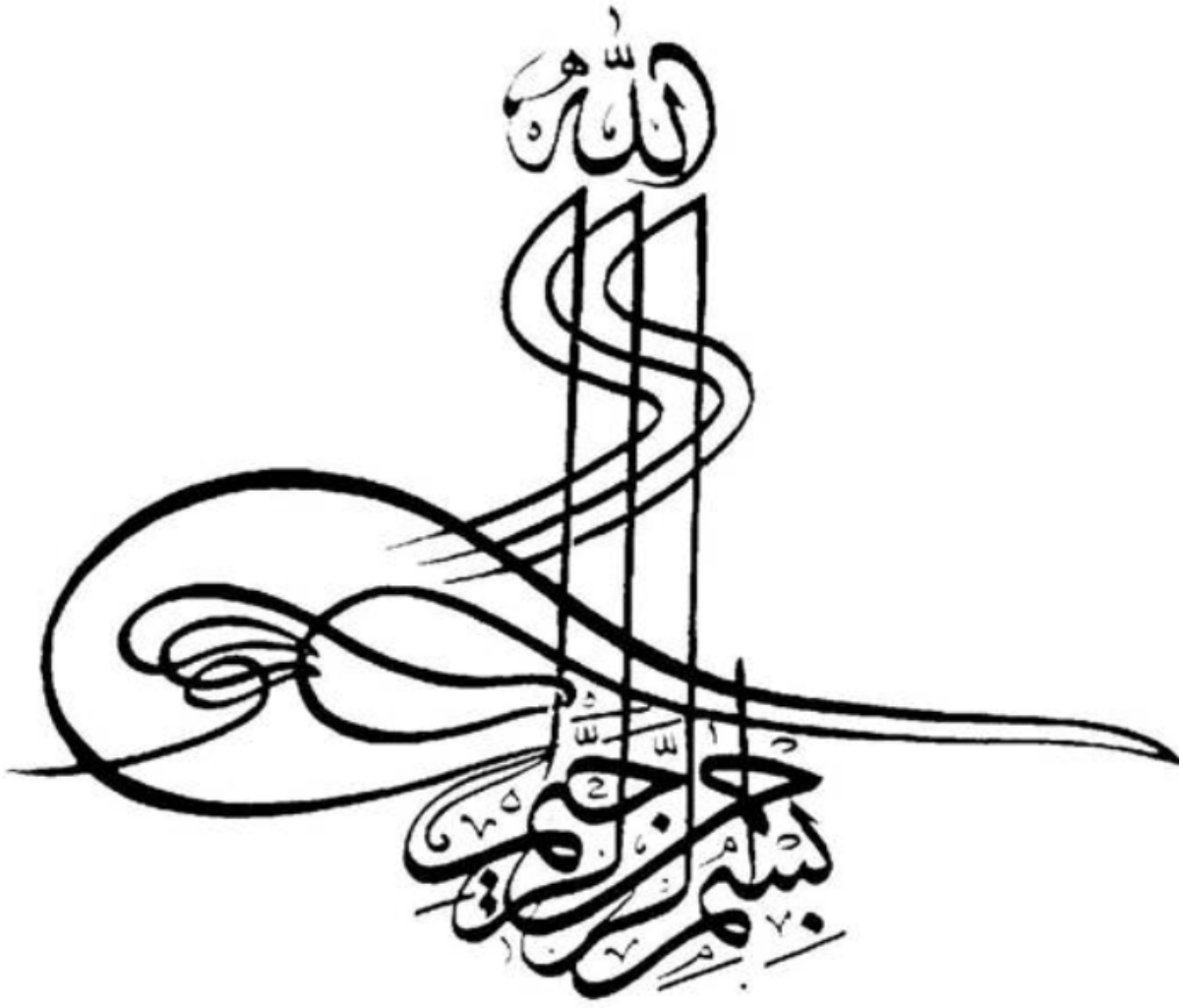
الله
بسم الله الرحمن الرحيم

دانشگاه تربیت مدرس

دانشکده علوم ریاضی

پایان نامه دوره کارشناسی ارشد آمار

استنباط بیزی برای مدل‌های آمیخته خطی چوله نرمال مستقل

توسط

مهناز احمدی

استاد راهنما

دکتر مجید جعفری خالدی

آذر ماه ۱۳۹۲

بسمه تعالی

دانشکده علوم ریاضی

تأییدیه اعضای هیأت داوران حاضر در جلسه دفاع از پایان‌نامه کارشناسی‌ارشد

اعضای هیأت داوران نسخهٔ نهایی پایان‌نامه **خانم مهناز احمدی** رشته **آمار** به شماره دانشجویی ۹۰۵۲۰۲۱۰۰۱ تحت عنوان: «**استنباط بیزی برای مدل‌های آمیخته خطی چوله نرمال مستقل**» را در تاریخ ۱۳۹۲/۹/۳۰ 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 مجید جعفری‌خالدی	استادیار	
۲- استاد ناظر داخلی	دکتر موسی گل‌علی‌زاده	استادیار	
۳- استاد ناظر خارجی	دکتر عادل محمدپور	دانشیار	
۴- استاد ناظر خارجی	دکتر زهرا رضایی	استادیار	
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 دانشگاه،دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می‌شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله)ی خود، مراتب را قبلا 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 ریاضی است که در سال ۱۳۹۲ در دانشکده علوم ریاضی دانشگاه تربیت مدرس به راهنمایی سرکار خانم/جناب آقای دکتر مجید جعفری خالکی، 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 مهناز احمدی دانشجوی رشته آمار ریاضی مقطع کارشناسی ارشد تعهد فوق وضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی: مهناز احمدی

تاریخ و امضا: ۹۲/۱۰/۱۷

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 هیات علمی، دانشجویان، دانش‌اموختگان 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 درامدهای حاصل از انها متعلق به دانشگاه می باشد ولی حقوق معنوی پدید اورندگان 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 دانش‌اموختگی 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 اثار 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 ائین 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 هیات 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 هیمان اجری دانشجوی رشته آبیاری ورودی سال تحصیلی ۱۳۹۵ مقطع کارشناسی ارشد دانشکده علوم کشاورزی متعهد می شوم کلیه نکات مندرج در ائین 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 ائین نامه فوق الاشعار به دانشگاه وکالت و نمایندگی می دهم که از طرف اینجانب نسبت به لغو امتیاز اختراع بنام بنده و یا هر گونه امتیاز دیگر و تغییر ان به نام دانشگاه اقدام نماید. ضمنا نسبت به جبران فوری ضرر و زیان حاصله بر اساس براورد دانشگاه اقدام خواهم نمود و بدینوسیله حق هر گونه اعتراض را از خود سلب نمودم»

امضا:

تاریخ: ۹۲/۱۰/۱۷

تقدیم به مقدسترین واژه ها در لغت نامه دلم

مادر مهربانم که زندگیم را مدیون مهر و عطوفت او میدانم

پدرم کوهی استوار و حامی من در تمام زندگی

همسرم که در سایه همیاری و همدلی او به این منظور نائل شدم

قدردانی

سپاس خدای را که سخنوران در ستودن او بمانند و شمارندگان شمردن نعمت‌های او ندانند و کوشندگان حق او را گزاردن نتوانند.

بدون شک جایگاه و منزلت معلم اجّل از آن است که در مقام قدردانی از زحمات بی‌شائبه او با زبان قاصر و دست ناتوان چیزی بنگارم. اما از آنجایی که تجلیل از معلم سپاس از انسانی است که هدف و غایت آفرینش را تأمین می‌کند و سلامت امانت‌هایی که به دستش سپرده‌اند، از تمامی اساتید بزرگوارم به ویژه اساتید محترم گروه آمار دانشگاه تربیت مدرس سپاسگزارم. بیش از همه از استاد با کمالات و شایسته، جناب آقای دکتر مجید جعفری خالدی که در کمال سعه صدر با حسن خلق و فروتنی از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این پایان‌نامه را بر عهده گرفتند کمال تشکر و قدردانی را دارم. از اساتید فرزانه، جناب آقای دکتر عادل محمدپور و سرکار خانم دکتر زهرا رضایی قهرودی که زحمت داوری این پایان‌نامه را متقبل شدند سپاسگزارم.

در پایان از پدر و مادر عزیزم و همسر مهربانم که سهم بسزایی در پیشرفت‌های زندگیم داشته‌اند و مسیر بهتر زیستن را به من آموخته‌اند صمیمانه تشکر می‌کنم.

مهناز احمدی

آذر ۱۳۹۲

چکیده

در بسیاری از مسائل آماری برای مدل‌بندی داده‌هایی که در طول زمان و به طور مکرر اندازه‌گیری می‌شوند، معمولاً مدل آمیخته خطی نرمال که در آن توزیع اثرات تصادفی و خطاها نرمال فرض می‌شود، به کار می‌رود. اما هنگامی که در میان داده‌ها، داده دور افتاده وجود داشته باشد، این مدل برازش خوبی به داده‌ها ندارد. برای حل این مشکل می‌توان به جای توزیع نرمال از کلاس توزیع‌های نرمال مستقل استفاده کرد، که دارای دم‌های کلفت‌تر از نرمال هستند و در حضور داده‌های دورافتاده عملکرد خوبی دارند. از طرف دیگر در مسائل کاربردی با موارد متعددی مواجه می‌شویم که علاوه بر حضور داده‌های دورافتاده، توزیع داده‌ها چوله است. در اینگونه مسائل می‌توان از توزیع‌های انعطاف پذیرتر چوله نرمال مستقل استفاده کرد که قادرند دو ویژگی چولگی و دم‌های کلفت را همزمان مدل کنند. در این پایان‌نامه، ضمن مطالعه مدل‌های آمیخته خطی چوله نرمال مستقل، برخی از خواص و ویژگی‌های این مدل‌ها بررسی می‌شوند. سپس مدل‌های آمیخته خطی چوله نرمال مستقل هنگامی که در میان داده‌ها، داده سانسور شده وجود دارد، مورد مطالعه قرار می‌گیرند. لازم به ذکر است استنباط این مدل‌ها به روش بیزی انجام می‌گیرد که در آن برای پارامترهای مدل توزیع پیشین اختیار شده و براساس روش‌های مونت کارلوی زنجیر مارکوفی از جمله الگوریتم گیبز از توزیع پسین نمونه‌گیری می‌شود. در انتها نحوه کاربست این مدل‌ها روی داده‌های روان‌آب به نمایش گذاشته و مدل مناسب‌تر از بین مدل‌های معرفی شده بر اساس معیار اطلاع انحراف انتخاب و تحلیل می‌شود.

واژه‌های کلیدی : مدل آمیخته خطی، دورافتاده، نرمال مستقل، چوله نرمال مستقل، روش بیزی، معیار اطلاع انحراف.

فهرست مندرجات

۱ مقدمات و مفاهیم مورد نیاز ۱

۱.۱ مقدمه ۱

۲.۱ مدل‌های آمیخته خطی ۴

۳.۱ معرفی توزیع‌های بیضوی ۶

۴.۱ توزیع چوله نرمال ۱۴

۵.۱ مدل‌های متغیر پنهان ۱۷

۱.۵.۱ الگوریتم EM ۲۱

۲.۵.۱ استنباط بیزی ۲۳

٦.١ معیارهای بیزی ۳۱

۲ تحلیل بیزی مدل‌های آمیخته خطی نرمال مستقل با استفاده از فرمول معکوس بیز ۳۷

۱.۲ مقدمه ۳۷

۲.۲ توزیع‌های نرمال مستقل چند متغیره ۳۸

۳.۲ مدل آمیخته خطی نرمال مستقل ۴۲

۴.۲ استنباط بیزی مدل آمیخته خطی نرمال مستقل ۴۵

۵.۲ مثال کاربردی : بررسی افزایش فاصله ارتودنسی با افزایش سن ۵۸

۳ مدل‌های آمیخته خطی چوله نرمال مستقل ٦٧

۱.۳ مقدمه ٦٧

۲.۳ خانواده توزیع‌های چوله نرمال مستقل چند متغیره ٦٨

۳.۳ مدل آمیخته خطی چوله نرمال مستقل ۷۵

۴.۳ استنباط بیزی مدل ۷۷

۵.۳ مدل آمیخته خطی چوله نرمال مستقل در حضور داده‌های سانسور شده ۸٦

۱.۵.۳ مروری بر انواع داده‌های سانسور شده ۸۷

۲.۵.۳ فرمول‌بندی مدل ۸۹

۳.۵.۳ لگاریتم تابع درستنمایی ۹۰

۴.۵.۳ استنباط بیزی مدل ۹۳

۴ مثال کاربردی ۹٦

۱.۴ مقدمه ۹٦

۲.۴ معرفی داده‌ها ۹۷

۳.۴ برازش مدل‌های آمیخته خطی به داده‌ها ۱۰۲

۴.۴ بحث و نتیجه‌گیری ۱۰۹

الف برنامه‌های WinBUGS **۱۲۱**

لیست اشکال

۱.۳.۱ تابع چگالی برخی از توزیع‌های شناخته شده بیضوی ۸

۲.۳.۱ نمودار کانتورهای توزیع‌های شناخته‌شده بیضوی ۹

۳.۳.۱ توزیع نرمال چند متغیره با ضریب همبستگی الف) $\rho = ۰/۷$ ب) $\rho = ۰$ . . . ۱۰

۴.۳.۱ هیستوگرام داده‌های شبیه سازی شده از متغیر تصادفی Y ۱۳

۵.۴.۱ تابع چگالی چوله نرمال به ازای مقادیر مختلف پارامتر چولگی λ ۱۴

۶.۵.۱ منحنی فراوانی داده‌های یک توزیع آمیخته ۱۸

۱.۵.۲ هیستوگرام و نمودارهای $Q-Q$ نرمال برای برآوردهای بیز تجربی اثرات تصادفی ۶۱

۲.۵.۲ هیستوگرام داده‌های ارتودنسی ۶۲

۳.۵.۲ نمودارهای شاخص $K(P, P_{(-i)})$ برای توزیع نرمال، توزیع تی استیودنت، توزیع اسلش، توزیع نرمال آلوده. ۶۵

۱.۲.۴ روان‌آب سطحی در دامنه کوه ۹۸

۲.۳.۴ (الف) نمودار $Q-Q$ نرمال و هیستوگرام برآورد بیز تجربی b_{i0}، (ب) نمودار $Q-Q$ نرمال و هیستوگرام برآورد بیز تجربی b_{i1} ۱۰۳

۳.۳.۴ هیستوگرام داده‌های روان‌آب ۱۰۴

فصل ۱

مقدمات و مفاهیم مورد نیاز

۱.۱ مقدمه

برای تحلیل داده‌هایی که در طول زمان و به طور مکرر اندازه‌گیری می‌شوند، اغلب از مدل‌های آمیخته خطی[۱] (LMM) (لیرد و وار، ۱۹۸۲) به دلیل انعطاف‌پذیری بالای آن‌ها برای در نظر گرفتن همبستگی درون گروهی و تغییرپذیری میان گروهی داده‌ها استفاده می‌شود. شناخته‌شده‌ترین مدل آمیخته خطی برای یک متغیر پاسخ پیوسته، مدل آمیخته خطی نرمال[۲] (NLMM) است، که در آن توزیع اثرات تصادفی و خطاهای درون گروهی نرمال فرض می‌شود. لذا این مدل نسبت به داده‌های دورافتاده حساس است. برای حل این مشکل به جای توزیع نرمال می‌توان از یک کلاس انعطاف‌پذیرتر از توزیع‌ها استفاده کرد. پین هیرو و همکاران (۲۰۰۱) یک مدل آمیخته خطی تی استیودنت[۳] (TLMM) پیشنهاد کردند، که در حضور داده‌های دورافتاده عملکرد خوبی دارد. رزا و

[۱] Linear Mixed Models

[۲] Normal Linear Mixed Model

[۳] Student's T Linear Mixed Model

همکاران (۲۰۰۳) نیز از یک زیرکلاسی از توزیع‌های بیضوی[۴] استفاده کردند، که توزیع‌های نرمال مستقل[۵] (NI) (لیو، ۱۹۹۶) نامیده می‌شوند و در یک چارچوب بیزی، استنباط مدل‌های آمیخته خطی نرمال مستقل را ارائه نمودند. اما حالتی را در نظر بگیرید، که چگالی داده‌ها چوله باشد، در این صورت فرض نرمال مستقل برای توزیع اثرات تصادفی و خطاها مناسب نیست. در این حالت، عموماً استفاده از توزیع چوله نرمال (آزالینی، ۱۹۸۵) به دلیل ویژگی‌های جالب آن مد نظر قرار می‌گیرد. جارا و همکاران (۲۰۰۸) مدل آمیخته خطی چوله نرمال[۶] (SNLMM) را معرفی نموده و به استنباط بیزی پارامترهای مدل پرداختند. در مسائل کاربردی با موارد متعددی مواجه می‌شویم، که علاوه بر اینکه توزیع داده‌ها نامتقارن است، داده دورافتاده نیز در میان آن‌ها وجود دارد. برانکو و دی (۲۰۰۱) یک کلاس جدید از توزیع‌های نیرومند تحت عنوان توزیع‌های آمیخته مقیاسی از چوله نرمال[۷] را توسعه دادند. این خانواده جدید از توزیع‌ها کاربرد گسترده‌ای در مدل‌های آماری دارند، زیرا قادرند دو ویژگی چولگی و دم‌های کلفت را همزمان مدل کنند. لچاس و همکاران (b و a ۲۰۰۹) و بندیاپادهیای و همکاران (۲۰۱۰) از یک زیرکلاسی از توزیع‌های آمیخته مقیاسی از چوله نرمال به نام توزیع‌های چوله نرمال مستقل[۸] (SNI) استفاده کردند و مدل آمیخته خطی چوله نرمال مستقل[۹] (SNILMM) را که در آن توزیع اثرات تصادفی SNI و توزیع خطاهاNI فرض می‌شود، پیشنهاد دادند. در ضمن لچاس و همکاران (۲۰۱۰) بر اساس روش ماکسیمم درستنمایی به برآورد پارامترهای مدل پرداختند. در ادامه بندیاپادهیای و همکاران (۲۰۱۲) استنباط بیزی مدل را در حضور داده‌های سانسور شده مورد بررسی

[۴] Elliptical
[۵] Normal Independent
[۶] Skew Normal Linear Mixed Model
[۷] Scale Mixture of Skew Normal Distributions
[۸] Skew Normal Independent
[۹] Skew Normal Independent Linear Mixed Model

قرار دادند. در تمام مقالاتی که در بالا اشاره شد برای انجام رهیافت بیزی از روش‌های مونت کارلوی زنجیر مارکوفی[۱۰] (MCMC) مبتنی بر الگوریتم نمونه‌گیری گیبز و الگوریتم متروپولیس هستینگس استفاده می‌شوند. روش‌های MCMC، به عنوان روش‌های مکرر با دو مسئله عمده وابستگی مارکوفی مقادیر تولید شده و مشکلات مربوط به همگرایی زنجیر مواجه هستند. برای حل این مسائل، تان و همکاران (۲۰۰۳ و ۲۰۱۰) روشی نامکرر بر اساس فرمول معکوس بیز[۱۱] (IBF) پیشنهاد دادند، که برای پیاده‌سازی آن از روش نمونه‌گیری بازنمونه‌گیری مهم[۱۲] (SIR) و الگوریتم امیدگیری و ماکسیمم سازی[۱۳] (EM) برای تعیین مد توزیع پسین استفاده می‌شود. بدین ترتیب می‌توان نمونه‌هایی مستقل از توزیع پسین تولید کرد، لذا با مشکلات مربوط به روش‌های MCMC مواجه نخواهیم بود. لچاس و همکاران (۲۰۱۲) روش نمونه‌گیری نامکرر مبتنی بر IBF را برای مدل‌های آمیخته خطی نرمال مستقل به کار بردند. همچنین وانگ و فان (۲۰۱۲) با ترکیب روش نمونه‌گیری گیبز و فرمول معکوس بیز، به تحلیل بیزی مدل‌های آمیخته خطی تی استیودنت چند متغیره پرداختند.

در این پایان‌نامه، ابتدا با خواص و ویژگی‌های مدل‌های آمیخته خطی نرمال مستقل و مدل‌های آمیخته خطی چوله نرمال مستقل آشنا شده، سپس برای تحلیل آن‌ها رهیافت بیزی مبتنی بر روش نمونه‌گیری فرمول معکوس بیز و روش‌های مونت کارلوی زنجیر مارکوفی به کار گرفته می‌شوند. در پایان کاربست روش بر مبنای تحلیل یک مجموعه داده واقعی نشان داده می‌شود. برای انجام این اهداف فصول پایان‌نامه به صورت زیر تدوین گردیده است. در ادامه این فصل تعاریف و مفاهیم مورد نیاز پایان‌نامه بیان می‌شود. در فصل دوم پس از معرفی مختصری از توزیع‌های نرمال مستقل، مدل‌های آمیخته

---

[۱۰] Markov Chain Monte Carlo Methods

[۱۱] Inverse Bayes Formula

[۱۲] Sampling Importance Resampling

[۱۳] Expectation Maximization

خطی نرمال مستقل و برخی نتایج استنباطی با استفاده از روش IBF ارائه می‌گردد. بعلاوه با یک مثال کاربردی، مقایسه مدل‌های آمیخته خطی مختلف با کمک معیارهای انتخاب مدل صورت می‌گیرد. در فصل سوم پس از معرفی مختصری از توزیع‌های چوله نرمال مستقل، مدل‌های آمیخته خطی چوله نرمال مستقل و استنباط بیزی این مدل‌ها بر اساس روش‌های مونت کارلوی زنجیر مارکوفی از جمله الگوریتم گیبز مورد بررسی قرار می‌گیرند. رهیافت برای هنگامی‌که داده‌ها به‌صورت دقیق مشاهده نشده و در بین آن‌ها داده‌های سانسور شده نیز وجود دارد، توسعه می‌یابد. در فصل چهارم نحوه کاربست مدل‌های آمیخته خطی چوله نرمال مستقل در تحلیل داده‌های روان آب بیان شده و مدل آمیخته خطی چوله اسلش تعمیم یافته با مدل‌های ارائه شده در فصل سوم مورد مقایسه قرار می‌گیرد.

۲.۱ مدل‌های آمیخته خطی

برای بیان یک مدل آمیخته خطی، ابتدا از یک مدل رگرسیونی شروع می‌کنیم. مدل رگرسیونی خطی به فرم

$$Y_i = X_i\beta + \varepsilon_i \qquad (۱.۲.۱)$$

نمایش داده می‌شود، که در آن معمولاً فرض می‌شود

$$\varepsilon_i \overset{iid.}{\sim} N(۰, \sigma^۲).$$

در معادله (۱.۲.۱) متغیر پاسخ Y_i توسط متغیر مستقل X_i توضیح داده می‌شود و β اثرات ثابت را نشان می‌دهد. در تحلیل بسیاری از داده‌ها از مدل رگرسیونی خطی استفاده می‌شود. اما در تحلیل داده‌هایی که در طول زمان به طور مکرر اندازه‌گیری می‌شوند، این مدل مناسب نیست، زیرا ممکن است در طول

زمان اندازه‌گیری، برخی از عوامل غیر قابل مشاهده روی متغیر پاسخ تأثیرگذار باشند، که در این مدل در نظر گرفته نمی‌شود. این عوامل تصادفی را متغیرهای پنهان می‌نامند و در بخش ۵.۱ به طور کامل معرفی می‌شوند. لحاظ نمودن این متغیرهای پنهان که اثرات تصادفی نامیده می‌شوند منجر به ایجاد مدلی تحت عنوان مدل آمیخته خطی می‌شود. بعبارت دیگر، مدل آمیخته خطی یک مدل آماری شامل اثرات آمیخته (ثابت و تصادفی) است. در تحلیل آماری بسیاری از داده‌های همبسته از جمله داده‌های حاصل از نمونه‌گیری خوشه‌ای، داده‌های طولی با اندازه‌گیری مکرر و داده‌های زوجی مانند داده‌های چشم مدل آمیخته خطی به طور گسترده مورد استفاده قرار می‌گیرد. در ادامه کاربرد این مدل در مطالعات طولی بطور مختصر توضیح داده می‌شود. در بسیاری از تحقیقات علوم پزشکی، پزشک برای بررسی تأثیر روش‌های درمانی مختلف، بیمار را در طول زمان و به طور مکرر مورد معاینه قرار می‌دهد. مشاهدات حاصل، اندازه‌گیری مکرر نام دارد. موقعیت‌هایی که اندازه‌گیری صورت می‌گیرد، لزوماً زمان نیستند. داده‌های مکرری که در آن موقعیت‌های تکرار مشاهدات نقاط زمانی هستند، داده‌های طولی و مطالعات از این نوع، مطالعات طولی نامیده می‌شود. هدف ابتدایی در یک مطالعه طولی بررسی تغییرات متغیر پاسخ در طول زمان و عوامل مؤثر بر آن است. زمانی که درباره یک مطالعه طولی صحبت می‌شود، دو دسته از تغییرات اهمیت پیدا می‌کنند، یکی تغییرات هر بیمار خاص به تنهایی در طول زمان که بررسی آنها تنها از طریق یک تحلیل طولی امکان‌پذیر بوده و دیگری تغییرات بین بیماران است. برای در نظر گرفتن هر دوی این تغییرات لزوم استفاده از یک مدل آمیخته خطی مشخص می‌شود.

مدل آمیخته خطی نرمال به فرم

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \qquad (۲.۲.۱)$$

نمایش داده می‌شود، که در آن فرض می‌شود

$$\begin{pmatrix} b_i \\ \varepsilon_i \end{pmatrix} \overset{ind.}{\sim} N_{n_i+q}\left(\begin{pmatrix} \mathbf{o} \\ \mathbf{o} \end{pmatrix}, \begin{pmatrix} D & \mathbf{o} \\ \mathbf{o} & \Sigma_i \end{pmatrix}\right)$$

در معادله(۲.۲.۱) Y_i بردار n_i بعدی پاسخ‌ها در گروه iام، $i = ۱, ..., n$، β بردار p بعدی ضرایب رگرسیونی(اثرات ثابت)، Z_i ماتریس طرح $n_i \times q$ برای اثرات تصادفی مشاهدات در گروه iام، X_i ماتریس طرح $n_i \times p$ برای اثرات ثابت مشاهدات در گروه iام و b_i بردار q بعدی اثرات تصادفی در گروه iام و همچنین ε_i بردار n_i بعدی خطاها در گروه iام است. D و Σ_i به ترتیب ماتریس‌های پراکندگی اثرات تصادفی و خطاها در گروه iام هستند. با توجه به این مفروضات توزیع متغیر پاسخ نرمال است. اما چنین فرضی ممکن است برقرار نباشد. یک راه حل استفاده از توزیع‌های انعطاف‌پذیر مانند توزیع‌های بیضوی و توزیع‌های چوله نرمال است، که در ادامه معرفی می‌شوند.

۳.۱ معرفی توزیع‌های بیضوی

بردار تصادفی $X = (X_۱, ..., X_n)^\top$ دارای توزیع بیضوی با بردار پارامترهای n بعدی μ، ماتریس مقیاس $n \times n$ بعدی همیشه مثبت Σ و تابع اسکالر ψ است، هرگاه تابع مشخصه آن به فرم

$$E[exp(it^\top X)] = exp(it^\top \mu).\psi(t^\top \Sigma t)$$

باشد، که در آن $t^\top = (t_۱, ..., t_n)$. ψ تابع مولد مشخصه X نامیده می‌شود. به طور مثال برای تابع مولد مشخصه $\psi(u) = exp(-u/۲)$، توزیع نرمال چند متغیره حاصل می‌شود. لازم به ذکر است هر تابع ψ نمی‌تواند به عنوان تابع مولد مشخصه استفاده شود. بدیهی است که باید $\psi(۰) = ۱$. یک شرط لازم و کافی برای اینکه ψ تابع مولد مشخصه توزیع بیضوی n بعدی باشد، در قضیه ۲.۲ فنگ و