

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

۸۷۶۱/۱۵
۸۸۱/۲۴



دانشگاه شهید بهشتی

دانشکده علوم ریاضی

رساله دکتری آمار

تحلیل پاسخ‌های همبسته‌ی آمیخته‌ی پیوسته و
ترتیبی با و بدون داده‌های گم‌شده

نگارنده

احسان بهرامی سامانی

استاد راهنما

دکتر مجتبی گنجعلی

استاد مشاور

دکتر احمد خدادادی

۱۳۸۷

۱۱۲۰۲۶



۱۳۸۸ / ۱ / ۲۱

کلیه حقوق اعم از چاپ و تکثیر، نسخه برداری، ترجمه، اقتباس و ... از این پایان نامه برای دانشگاه شهید بهشتی محفوظ است. نقل مطالب با ذکر مأخذ بلامانع است.

تقدیم به
پیشگاه لایزال یگانه اش

تقدیم به همسر مهربانم

و همه کسانی که دوستشان دارم.

قدردانی و تشکر

خداوندا، سپاس ترا بر آنکه پرده‌ی تاریکی شب را به نور صبح شکافتی و ما را از روشنایی روز بهره‌مند ساختی و به منافع روزی‌ها بینا فرمودی.

پروردگارا، به من توفیق بده تا همواره سخن حق را بگویم، هرچند دشوار باشد و هرگز کورکورانه در علم از روی بی‌اطلاعی سخن نگویم. معبودا، قوی‌ترین نیروهایت را در من به هنگام خستگی قرار ده و مرا به سستی در عبادت و ناپینایی در تشخیص راه تو و انجام عمل خلاف دوستی تو و پیوستن با کسی که از تو جدا شود و جدا شدن از کسی که با تو به پیوندد، دچار نساز.

از استاد ارجمندم جناب آقای دکتر مجتبی گنجعلی که زحمت راهنمایی این رساله را پذیرفته و بدون شک انجام مراحل مختلف این رساله بدون راهنمایی‌های گرانبها و تلاش‌های ارزنده‌ی ایشان ممکن نبود، کمال تشکر و قدردانی را دارم.

از استاد گرامی جناب آقای دکتر احمد خدادادی که مشاوره‌ی این رساله را بر عهده داشتند صمیمانه تشکر و قدردانی می‌نمایم.

از اساتید محترم جناب آقای دکتر کاظمی، دکتر گرامی، دکتر خزایی و دکتر فرید روحانی که زحمت مطالعه و داوری این رساله را تقبل فرمودند کمال تشکر را دارم.

با تشکر و سپاس فراوان از همه‌ی دوستان و همکلاسی‌هایم به ویژه خانم سمانه افتخاری، خانم زهرا رضایی، آقای جواد اطمینان، علیرضا طاهریون، هادی پورمحمدی و امین شرقی ایدو که کمک‌های زیادی به من نمودند.

چکیده

در این رساله برخی مدل‌های توأم یک متغیره، چند متغیره و چند متغیره‌ی طولی با پاسخ‌های همبسته‌ی آمیخته‌ی پیوسته و ترتیبی با و بدون داده‌های گم‌شده معرفی شده‌اند. یک شیوه بر مبنای تابع درست‌نمایی به‌دست آمده از مدل، برای به‌دست آوردن برآورد پارامترهای هر مدل استفاده شده است. برخی مانده‌های پی‌یرسونی تصحیح شده برای هریک از مدل‌ها نیز معرفی شده‌اند که در آن‌ها همبستگی بین پاسخ‌ها، و در حالت گم‌شدگی داده‌ها، مکانیسم گم‌شدگی نیز در نظر گرفته شده است. برای مدل متغیر پنهان آمیخته‌ی تک متغیره، مقایسه‌ای میان ضریب همبستگی پی‌یرسون و ضریب همبستگی به‌دست آورده شده توسط مدل، با یک مطالعه‌ی شبیه‌سازی انجام شده است. برای مدل همزمان تک متغیره‌ی آمیخته با مقادیر گم‌شده‌ی غیر تصادفی بالقوه، تحلیل حساسیت برای بررسی تأثیر فرض موجود در مکانیسم گم‌شدگی، روی استنباط در مورد پارامترهای مدل انجام شده است. برای تشریح سودمندی مدل‌های پیشنهاد شده برخی داده‌های واقعی تحلیل شده‌اند.

کلمات کلیدی: مدل همزمان، پاسخ‌های آمیخته‌ی پیوسته و ترتیبی، تحلیل حساسیت،

مطالعات چند متغیره‌ی طولی، پاسخ‌های آمیخته‌ی گم‌شده.

پیشگفتار

یکی از مسائل مهم مورد توجه پژوهشگران آمار کاربردی پیوند بین چند متغیر پیوسته و رسته‌ای است. عمومی‌ترین ابزار آماری برای این‌گونه مسائل استفاده از رگرسیون و همبستگی است. همبستگی پی‌یرسون تنها پیوند خطی را اندازه‌گیری می‌کند که می‌توان این را برای انواع متغیرهای پیوسته و گسسته و نیز آمیخته‌ای از آن متغیرها مورد بررسی قرار داد. از آنجایی که داده‌های آمیخته‌ی پیوسته و گسسته کاربردهای بسیار زیادی در علوم مختلف مانند پزشکی و ... دارند، از این جهت بررسی پیوند بین متغیرهای پیوسته و گسسته در داده‌های آمیخته و مدل بندی رگرسیونی آن‌ها از اهمیت بالایی برخوردار است.

در بررسی پیوند بین برداری از متغیرهای پیوسته و برداری از متغیرهای گسسته از نوع رسته‌ای مدلی تحت عنوان مدل مکانی عام مطرح شد. در این مدل متغیرهای رسته‌ای یک جدول پیشابندی، به وسیله‌ی تقاطع سطوح‌شان، ایجاد می‌کنند که در هر خانه فرض می‌شود متغیرهای پیوسته دارای توزیع نرمال چند متغیره هستند. مسأله‌ای که در این مدل‌ها مورد توجه قرار نگرفته است امکان ترتیبی بودن سطح‌های متغیر رسته‌ای است. در واقع مسأله‌ی اصلی پیوند بین یک متغیر پیوسته و یک متغیر ترتیبی و همچنین تأثیر متغیرهای تبیینی روی این متغیرها است که از اهمیت کاربردی زیادی برخوردار است. تعمیم این موضوع به حالت

چند متغیره (برداری از متغیرهای پیوسته و برداری از متغیرهای ترتیبی) نیز از اهمیت زیادی برخوردار است. استفاده از این مدل در حالت پاسخ‌های گم‌شده‌ی قابل چشم‌پوشی یا غیر قابل چشم‌پوشی نیز اهمیت دارد که مورد بررسی قرار خواهد گرفت.

اولین بارتیت (۱۹۵۴، ۱۹۵۵ a و ۱۹۵۵ b) به بررسی پیوند بین یک متغیر پیوسته و یک متغیر گسسته‌ی دودویی پرداخت. برای این منظور دو متغیر پیوسته‌ی Z و Y^* را در نظر گرفت و فرض کرد ضریب همبستگی این دو متغیر در جامعه ρ باشد. سپس در این زمینه سه مدل ارائه نمود. در مدل اول که به عنوان یک مدل بنیادی، پیوند بین دو متغیر پیوسته‌ی Z و Y^* را مورد بررسی قرار می‌داد، Z و Y^* دارای توزیع نرمال دو متغیره بودند. با در نظر گرفتن نمونه‌ی تصادفی (Z_i, Y_i^*) ، $i = 1, \dots, n$ ، برآورد ρ را که ضریب همبستگی گشتاوری حاصل ضریبی پی‌یرسون (۱۹۰۹) یا ضریب همبستگی معمولی موسوم است به دست آورد. در مدل دوم گاهی مقادیر Y^* مشاهده نمی‌شوند ولی مقادیر تبدیل شده‌ی این متغیر به صورت دودویی، Y_i ، مشاهده می‌شوند. در این مدل نیز تیت با در نظر گرفتن نمونه‌ی تصادفی n تایی از (Z_i, Y_i) ، $i = 1, \dots, n$ ، برآورد ρ را، که در مدل اول ظاهر شده است، معرفی کرد و این برآورد را ضریب همبستگی دورشته‌ای نامید. در مدل سوم در مورد Y^* یا توزیع آن هیچ فرضی در نظر گرفته نشده است. توزیع Y را، توزیع برنولی با پارامتر p در نظر گرفت و فرض کرد که توزیع شرطی Z به شرط $Y = y$ ، نرمال با میانگین μ_y و واریانس σ^2 است. در این مدل برآورد ρ به ضریب همبستگی دورشته‌ای نقطه‌ای موسوم است.

اولکین و تیت (۱۹۶۱) حالت‌هایی را که تیت به آن‌ها اشاره کرده بود تعمیم دادند و مدلی را تحت عنوان مدل مکانی عام مطرح نمودند. در این مدل پیوند بین برداری از متغیرهای رسته‌ای و برداری از متغیرهای پیوسته مورد بررسی قرار گرفت. در این مدل

متغیرهای رسته‌ای جدول پیشابندی به وسیله‌ی تقاطع سطوح‌شان ایجاد می‌کنند که در هر خانه فرض می‌شود متغیرهای پیوسته دارای توزیع نرمال چند متغیره هستند.

کاکس (۱۹۷۲) با استفاده از تجزیه‌ی توزیع توأم متغیرهای پیوسته و گسسته به توزیع حاشیه‌ای متغیر پیوسته و توزیع شرطی متغیر گسسته به شرط متغیر پیوسته، مدلی را ارائه نمود که پیوند بین متغیرهای پیوسته و گسسته را مورد بررسی قرار می‌داد.

هکمن (۱۹۷۸) مدلی تحت عنوان مدل همزمان برای پاسخ‌های آمیخته‌ی پیوسته و گسسته ارائه کرد. در این مدل با استفاده از همبستگی بین خطاهایی که در مدل مورد نظر قرار داشت، به بررسی پیوند بین پاسخ‌های آمیخته‌ی پیوسته و گسسته پرداخت. کاتالانو و ریان (۱۹۹۲) این مدل را برای داده‌های خوشه‌ای (برای مثال داده‌های مربوط به میزان درآمد اعضای خانواده‌ی یکسال در کل خانوارهای استان تهران که در آن خوشه عبارت است از یک خانوار) تعمیم دادند.

کاکس و ویرموتس (۱۹۹۲) به مقایسه‌ی مدل مکانی عام و مدلی که کاکس در سال ۱۹۷۲ ارائه نموده بود، پرداختند. مفهوم گم‌شدگی داده‌ها توسط رابین (۱۹۷۶) بیان شد و سپس لیتل و شالتر (۱۹۸۵) برآورد ماکسیمم درست‌نمایی را برای مدل‌های آمیخته با پاسخ‌های پیوسته و اسمی که دارای گم‌شدگی باشند به دست آوردند. فیتز موریس و لیرد (۱۹۹۵) مدل رگرسیونی را برای داده‌های آمیخته‌ی پیوسته و گسسته خوشه‌ای ارائه نمودند. فیتز موریس و لیرد (۱۹۹۷) این مدل را برای حالتی که برخی از پاسخ‌ها دارای گم‌شدگی هستند به کار بردند.

دیگل و همکاران (۱۹۹۴) مدل‌هایی را روی داده‌های طولی آمیخته مطرح کردند. داده‌های طولی آمیخته داده‌های هستند که در آن‌ها پاسخ‌های آمیخته برای هر آزمودنی در

طول زمان تکرار شده‌اند. از آنجایی که بین پاسخ‌ها برای هر آزمودنی همبستگی وجود دارد به روش‌های خاصی برای تحلیل آماری نیاز است تا بتوان استنباط‌های معتبری به دست آورد. برای این منظور مدل‌هایی چون مدل‌های حاشیه‌ای، اثرهای تصادفی و انتقالی معرفی نمودند. در مدل‌های حاشیه‌ای تأثیر متغیرهای تبیینی بر پاسخ‌ها به طور مجزا از همبستگی بین پاسخ‌ها برای یک آزمودنی معین مدل بندی می‌شود. همچنین در مدل‌های اثر تصادفی فرض می‌شود که برخی ضرایب رگرسیونی برای آزمودنی‌های مختلف به دلایل متفاوت به خود آزمودنی‌ها وابسته است و در مدل‌های انتقالی فرض بر این است که متغیر پاسخ در زمان t به متغیرهای پاسخ در زمان‌های قبلی وابسته می‌باشد. استفاده از مدل‌های حاشیه‌ای در داده‌های آمیخته توسط کاتالانو و ریان (۱۹۹۲) و بیکر (۱۹۹۵) صورت گرفت. لیتل و رابین (۲۰۰۲)، با استفاده از مفاهیم و مکانیسم‌های گم‌شدگی تابع درست‌نمایی را برای داده‌های آمیخته‌ی پیوسته و گسسته که دارای گم‌شدگی هستند، بیان نمودند. گنجعلی (۲۰۰۳) به بررسی پیوند بین یک متغیر پاسخ پیوسته و یک متغیر پاسخ گسسته‌ی دودویی به طوری که برخی از پاسخ‌ها گم‌شدگی غیرقابل چشم‌پوشی دارند، پرداخت و تأثیر متغیرهای تبیینی روی این دو پاسخ را به طور همزمان مدل بندی کرد.

ابراهیم و همکاران (۲۰۰۱) به بررسی پاسخ‌های گم‌شده در مدل‌های خطی آمیخته که مکانیسم گم‌شدگی آن‌ها غیرقابل چشم‌پوشی باشد، پرداختند. زارع و همکاران (۲۰۰۳) مدل‌های آمیخته‌ی با متغیر پنهان در داده‌های گم‌شده را مورد بررسی قرار دادند. لی و سانگ (۲۰۰۴) به تحلیل ماکسیمم تابع درست‌نمایی روی مدل‌هایی با متغیر پنهان در داده‌های آمیخته‌ی سلسله مراتبی پرداختند. گنجعلی و شفیع (۲۰۰۶) مدلی انتقالی برای پاسخ‌های آمیخته‌ی گسسته و پیوسته در داده‌های مرتب شده‌ی خوشه‌ای با پاسخ‌های گم‌شده ارائه

نمودند.

اسپیس (۲۰۰۶) با استفاده از معادلات برآوردگر تعمیم یافته و با فرض گم شدن قابل چشم‌پوشی مدلی را برای داده‌های آمیخته‌ی پیوسته و ترتیبی معرفی کرد. وی از معادلات درست‌نمایی برای برآورد پارامترها استفاده نکرد. از آن‌جا که اسپیس (۲۰۰۶) از جانهی چندگانه استفاده کرده است، در صورتی که داده‌ها از نوع گم شدن غیر قابل چشم‌پوشی باشند، برآوردهای پیشنهادی او اریب اند. لیون و کریر (۲۰۰۷) مدل مکانی عام را برای داده‌های آمیخته‌ی پیوسته و ترتیبی به کار بردند.

در این رساله، مدل همزمان جدیدی مشابه با مدل‌های هکمن (۱۹۷۸) معرفی می‌شود. پاسخ‌های این مدل متغیرهای پیوسته و ترتیبی همبسته‌اند و ممکن است دارای گم‌شدگی نیز باشند. همچنین تعمیمی از این مدل برای پاسخ‌های چند متغیره ارائه شده است. علاوه بر آن، با تعمیم مدل مربوط به پاسخ‌های طولی چند متغیره با داده‌های گم‌شده، مدل جدید و کاملی معرفی می‌شود. برنامه‌های کامپیوتری این مدل‌ها نیز به صورت بسته‌های جدیدی در نرم‌افزار R ، تنظیم شده است. همچنین روشی جدید برای تحلیل حساسیت نتایج به نوع گم‌شدگی پاسخ‌ها، معرفی خواهد شد.

در فصل اول این رساله، به تعاریف اساسی مانند مفهوم متغیر، مقیاس و انواع آن‌ها و همچنین مسأله‌ی گم‌شدگی و مکانیسم‌های آن و تابع درست‌نمایی داده‌ها و مفاهیم تحلیل حساسیت روی داده‌ها پرداخته شده است. در فصل دوم، مدل مربوط به پاسخ‌های تک متغیره‌ی همبسته‌ی آمیخته‌ی پیوسته و ترتیبی با و بدون داده‌های گم‌شده معرفی می‌شود. در فصل سوم مدل مربوط به پاسخ‌های چند متغیره‌ی همبسته‌ی آمیخته‌ی پیوسته و ترتیبی برای داده‌های کامل بیان شده است. در فصل چهارم، مدل مربوط به پاسخ‌های طولی چند متغیره‌ی

همبسته‌ی آمیخته با و بدون داده‌های گم‌شده معرفی گردیده است. در فصل پنجم، کاربرد مدل‌های بیان شده روی داده‌های واقعی را ارائه شده است.

فهرست مندرجات

۱	معرفی انواع متغیرها، مفاهیم گم‌شدگی و تحلیل حساسیت	۱
۱ مقدمه	۱.۱
۲ مقیاس، مقیاس‌سازی و متغیر	۲.۱
۳ انواع متغیرها	۳.۱
۷ مطالعات طولی	۴.۱
۷ مسئله‌ی گم‌شدگی در داده‌هایی با متغیرهای اسمی، پیوسته و ترتیبی	۵.۱
۱۱ تابع درست‌نمایی مربوط به داده‌های مشاهده شده	۶.۱

۱۷	تحلیل تأثیر موضعی	۷.۱
۲۰	منحنی‌های فضایی	۱.۷.۱
۲۳	محاسبه‌ی خمیدگی روبه‌ی هندسی	۲.۷.۱

۲ تحلیل پاسخ‌های همبسته‌ی آمیخته‌ی تک متغیره‌ی پیوسته و

ترتیبی، با و بدون داده‌های گم‌شده

۲۶	مقدمه	۱.۲
۲۷	تحلیل پاسخ‌های ترتیبی	۲.۲
۲۹	تحلیل پاسخ‌های پیوسته	۳.۲
۳۰	تحلیل پاسخ‌های همبسته‌ی آمیخته‌ی پیوسته و ترتیبی	۴.۲
		مدل مقطعی با پاسخ‌های تک متغیره‌ی همبسته‌ی	۱.۴.۲
	۳۱	آمیخته‌ی پیوسته و ترتیبی
۳۲	تابع درست‌نمایی	۲.۴.۲
۳۳	مانده‌های مدل	۳.۴.۲

۳۵	تحلیل پاسخ‌های همبسته‌ی آمیخته‌ی پیوسته و ترتیبی با پاسخ‌های گم‌شده	۵.۲
	مدل مقطعی با پاسخ‌های تک‌متغیره‌ی همبسته‌ی	۱.۵.۲
	آمیخته‌ی پیوسته و ترتیبی با داده‌های گم‌شده	۳۶
۳۸	تابع درست‌نمایی	۲.۵.۲
۴۱	مانده‌های مدل	۳.۵.۲

۳ تحلیل پاسخ‌های همبسته‌ی آمیخته‌ی چند متغیره‌ی پیوسته و

ترتیبی با و بدون داده‌های گم‌شده

۴۵

۴۵ مقدمه

۴۷ معرفی مدل مکانی عام

۵۰ مدلی جدید برای داده‌های همبسته‌ی آمیخته‌ی پیوسته و ترتیبی

۵۱ تابع درست‌نمایی

۵۴ تقریبی از تابع درست‌نمایی

۵۶ مانده‌های مدل

۴.۳ تحلیل پاسخ‌های چند متغیره‌ی همبسته‌ی آمیخته‌ی پیوسته و ترتیبی با

داده‌های گم‌شده

۵۸

۱.۴.۳	مدل مقطعی با پاسخ‌های چند متغیره‌ی همبسته‌ی
۵۸	آمیخته‌ی ترتیبی و پیوسته با داده‌های گم‌شده
۶۱	۲.۴.۳ تابع درست‌نمایی
۶۶	۳.۴.۳ مانده‌های مدل

۴ تحلیل پاسخ‌های طولی چند متغیره‌ی همبسته‌ی آمیخته‌ی

۶۸	پیوسته و ترتیبی با و بدون داده‌های گم‌شده
----	---

۶۸	۱.۴ مقدمه
۶۹	۲.۴ مدل اثرهای تصادفی

۳.۴ تحلیل پاسخ‌های چند متغیره‌ی طولی همبسته‌ی آمیخته‌ی پیوسته و

۷۲	ترتیبی با اثرهای تصادفی
۱.۳.۴	مدل طولی با اثرهای تصادفی برای پاسخ‌های چند
۷۲	متغیره‌ی همبسته‌ی آمیخته‌ی پیوسته و ترتیبی
۷۴	۲.۳.۴ تابع درست‌نمایی

۴.۴ تحلیل پاسخ‌های چند متغیره‌ی طولی همبسته‌ی آمیخته‌ی پیوسته و

۷۴	ترتیبی با داده‌های گم‌شده
----	---------------------------

	۱.۴.۴	مدل طولی با اثرهای تصادفی برای پاسخ‌های چند
		متغیره‌ی همبسته‌ی آمیخته‌ی پیوسته و ترتیبی با داده‌های گم‌شده . . . ۷۵
۷۷	۲.۴.۴	تابع درست‌نمایی
۸۳	۵.۴	شیوه‌ی دیگر مدل‌بندی
۸۵	۱.۵.۴	تابع درست‌نمایی
۸۶		۵ کاربرد
۸۶	۱.۵	مقدمه
	۲.۵	مدل آمیخته‌ی مقطعی: یک مطالعه‌ی شبیه‌سازی و کاربرد در
۸۷		داده‌های کبد چرب و میزان چاقی
۸۷	۱.۲.۵	مطالعه‌ی شبیه‌سازی
۸۹	۲.۲.۵	نتایج مربوط به شبیه‌سازی
۹۱	۳.۲.۵	معرفی داده‌ی کبد چرب و میزان چاقی
۹۴	۴.۲.۵	مدل مربوط به داده‌های کبد چرب و میزان چاقی
۹۵	۵.۲.۵	نتایج مربوط به مدل

۳.۵	مدل آمیخته‌ی مقطعی با پاسخ‌های گم‌شده: کاربرد در فراگیری
۹۶	زبان‌های خارجی
۹۷	۱.۳.۵ معرفی داده‌های فراگیری زبان‌های خارجی
۹۸	۲.۳.۵ مدل مربوط به داده‌های فراگیری زبان‌های خارجی
۹۹	۳.۳.۵ نتایج مربوط به مدل
۱۰۲	۴.۳.۵ تحلیل حساسیت
۴.۵	مدل آمیخته‌ی مقطعی: کاربرد در پوکی استخوان زنان
۱۰۷	۱.۴.۵ معرفی داده‌های پوکی استخوان زنان
۱۰۸	۲.۴.۵ مدل مربوط به داده‌های پوکی استخوان زنان
۱۰۹	۳.۴.۵ نتایج مربوط به مدل
۵.۵	مدل چند متغیره‌ی مقطعی: کاربرد در داده‌های کیفیت زندگی
۱۱۱	۱.۵.۵ معرفی داده‌های کیفیت زندگی
۱۱۳	۲.۵.۵ مدل مربوط به داده‌های کیفیت زندگی
۱۱۴	۳.۵.۵ نتایج مدل
۶.۵	مدل چند متغیره‌ی طولی: کاربرد در داده‌های ناهنجاری کودکان
۱۱۷	۱.۶.۵ معرفی داده‌های ناهنجاری کودکان
۱۱۸	۲.۶.۵ مدل مربوط به داده‌های ناهنجاری کودکان

۱۱۹ تابع درستمایی	۳.۶.۵
۱۲۰ نتایج مدل	۴.۶.۵
۱۲۲ نتیجه گیری و پیشنهادهای	۷.۵
۱۲۳	A واژه‌نامه‌ی فارسی به انگلیسی
۱۲۵	مراجع

فصل ۱

معرفی انواع متغیرها، مفاهیم گم‌شدگی و تحلیل حساسیت

۱.۱ مقدمه

در این فصل مفاهیم و تعاریف اساسی بیان می‌شوند که در تحلیل پاسخ‌های همبسته‌ی آمیخته‌ی پیوسته و ترتیبی مورد استفاده قرار می‌گیرند. این مفاهیم شامل معرفی انواع مقیاس‌های اندازه‌گیری، متغیرهای مورد نیاز و انواع آن‌ها، مفاهیم گم‌شدگی یا بی‌پاسخی داده‌ها، انواع مکانیسم‌های گم‌شدگی و چگونگی به‌دست آوردن تابع درست‌نمایی داده‌ها با و