

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

دانشگاه یزد

دانشکده ریاضی

گروه آمار

پایان نامه

برای دریافت درجه کارشناسی ارشد
آمار ریاضی

مدل رگرسیون وایبول نمایی شده‌ی لگاریتمی برای داده‌های
سانسور شده بازه‌ای

استاد راهنما: دکتر حمزه ترابی

استاد مشاور: دکتر علی دولتی

پژوهش و نگارش: سیدمحمد میرزایی

تیر ماه ۱۳۹۱

تقدیم به پدر و مادر عزیزم که بهترین مشوق من در تحصیل علم بوده‌اند.
سپاسگزارم از استاد راهنمای گرامی، دکتر حمزه ترابی که بدون حضور ایشان
پیمودن این راه برایم غیر ممکن بود.

سپاسگزارم از استاد مشاور گرامی، دکتر علی دولتی که در این راه یاریم
نموده‌اند.

سپاسگزارم از استادان گرامی، دکتر اکبر اصغر زاده و دکتر علی اکبر
جعفری که داوری این پایان‌نامه را به عهده گرفتند.

چکیده

در برخی از پژوهش‌های تحلیل بقاء، به دلیل وجود عوامل بازدارنده مانند زمان و هزینه، واحدهای نمونه به طور دقیق مشاهده نمی‌شوند، بلکه درون بازه‌های زمانی ویژه‌ای رخ می‌دهند؛ به این نوع سانسور، سانسور بازه‌ای گفته می‌شود. در این پایان‌نامه یک مدل رگرسیون مکان-مقیاس بر پایه توزیع وایبول نمایی شده، برای مدل‌بندی داده‌های سانسور شده بازه‌ای طول عمر در نظر گرفته می‌شود. این مدل، مدل رگرسیون وایبول نمایی شده لگاریتمی نامیده می‌شود و خانواده‌ای از مدل‌های رگرسیون پارامتری است که خود شامل مدل‌های دیگری است که کاربرد وسیعی در تحلیل داده‌های طول عمر دارند. برای برآورد پارامترهای مدل از روش‌های برآوردگر ماکسیم درست‌نمایی، برآوردگر جک‌نایف، بوت‌استرپ پارامتری و برآورد بیزی استفاده می‌شود. برای ارزیابی حساسیت مدل از دو روش تأثیر فراموضعی و تأثیر موضعی استفاده می‌شود. در روش تأثیر موضعی، ماتریس‌های مربوط به طرح‌های مختلف اغتشاش محاسبه شده است. مطالعات شبیه‌سازی نشان می‌دهد که باقی‌مانده انحراف تعدیل‌یافته برای این مدل، به طور تقریبی دارای توزیع نرمال استاندارد است و در نتیجه می‌توان از آن به عنوان معیاری برای سنجش اعتبار مدل استفاده کرد. این مدل به یک نمونه واقعی برازش داده و نتایج حاصل گزارش شده است.

فهرست مطالب

۱	مفاهیم و تعاریف اولیه	۱
۲	۱.۱ داده‌های سانسور شده	۱.۱
۲	۱.۱.۱ سانسور نوع اول	۱.۱.۱
۳	۲.۱.۱ سانسور نوع دوم	۲.۱.۱
۳	۳.۱.۱ سانسور راست و چپ	۳.۱.۱
۴	۴.۱.۱ سانسور تصادفی	۴.۱.۱
۶	۵.۱.۱ سانسور بازه‌ای	۵.۱.۱
۷	۲.۱ مباحث استنباط آماری و قضایای ریاضی مورد نیاز	۲.۱
۷	۱.۲.۱ قضیه حد مرکزی	۱.۲.۱
۸	۲.۲.۱ شبیه‌سازی	۲.۲.۱
۹	۳.۲.۱ زنجیر مارکوف پیوسته	۳.۲.۱
۹	۴.۲.۱ الگوریتم متروپولیس-هستینگس	۴.۲.۱
۱۱	۵.۲.۱ الگوریتم نمونه‌گیری گیبس	۵.۲.۱
۱۳	۶.۲.۱ انتخاب بهترین برآوردگر	۶.۲.۱
۱۴	۷.۲.۱ بسط تیلور	۷.۲.۱
۱۵	۸.۲.۱ معرفی عملگر O	۸.۲.۱
۱۷	۲ تحلیل بقاء	۲
۱۸	۱.۲ توزیع آماری متغیر طول عمر	۱.۲

۱۹	مدل‌های پارامتری	۱.۱.۲
۲۱	مدل‌های ناپارامتری	۲.۱.۲
۲۲	تحلیل رگرسیون	۲.۲
۲۲	مدل رگرسیون با نرخ شکست متناسب	۱.۲.۲
۲۴	مدل رگرسیون زمان شکست شتابیده	۲.۲.۲
۲۷	مدل رگرسیون وایبول نمایی شده‌ی لگاریتمی برای داده‌های سانسور شده بازه‌ای	۳
۲۸	مدل رگرسیون مکان-مقیاس بر پایه وایبول نمایی شده	۱.۳
۳۵	برآورد پارامترهای مدل رگرسیون	۲.۳
۳۵	برآوردگر ماکسیمم درست‌نمایی	۱.۲.۳
۳۶	برآوردگر جک‌نایف	۲.۲.۳
۳۸	روش باز نمونه‌گیری بوت‌استرپ	۳.۲.۳
۴۱	تحلیل بیزی	۴.۲.۳
۴۶	تحلیل حساسیت	۳.۳
۴۶	تأثیر فراموضعی	۱.۳.۳
۴۹	تأثیر موضعی	۲.۳.۳
۶۳	تحلیل باقی‌مانده‌ها	۴.۳
۶۳	باقی‌مانده‌های متداول در تحلیل بقاء	۱.۴.۳
۶۷	شبیه‌سازی	۲.۴.۳
۷۷		۴ کاربرد
۷۸	برآوردهای ماکسیمم درست‌نمایی و جک‌نایف	۱.۴
۷۹	برآوردهای روش بوت‌استرپ	۲.۴
۸۰	روش بیزی برای برآورد پارامترها	۳.۴
۸۳	تحلیل حساسیت مدل	۴.۴
۸۴	تحلیل باقی‌مانده‌ها	۵.۴

۸۶	۶.۴	بررسی اثر داده‌های مؤثر
۸۷	۷.۴	نیکویی برازش
۸۹			الف داده‌های واقعی
۹۱			ب برنامه‌های کامپیوتری
۹۱	۲.۴.۳	۱.ب برنامه‌های شبیه‌سازی بخش
۹۶	۴	۲.ب برآورد پارامترهای مدل فصل
۱۰۹	۴	۳.ب تحلیل حساسیت مدل فصل
۱۱۵	۴	۴.ب تحلیل باقی‌مانده‌های مدل فصل
۱۱۷			پ واژه‌نامه‌ی فارسی به انگلیسی
۱۲۱			ت واژه‌نامه‌ی انگلیسی به فارسی
۱۲۴			منابع و مراجع

فهرست شکل‌ها

- ۱۰۳ نمودار تابع چگالی وایبول نمایی شده برای مقادیر مختلف α با $\lambda = 2$ و $\gamma = 2$. ۳۰
- ۲۰۳ نمودار تابع چگالی وایبول نمایی شده برای مقادیر مختلف λ با $\alpha = 2/5$ و $\gamma = 2$. ۳۱
- ۳۰۳ نمودار تابع چگالی وایبول نمایی شده لگاریتمی استاندارد برای مقادیر مختلف λ . ۳۲
- ۴۰۳ نمودار چندک-چندک نرمال استاندارد برای باقی مانده‌های r_{Di}^* و r_{Mi}^*, r_{Ci}^*
 $n = 50, n = 100, n = 200$ ، درصد داده‌های سانسور از راست برابر 0° و
 $\lambda = 0/3$. ۷۰
- ۵۰۳ نمودار چندک-چندک نرمال استاندارد برای باقی مانده‌های r_{Di}^* و r_{Mi}^*, r_{Ci}^*
 $n = 50, n = 100, n = 200$ ، درصد داده‌های سانسور از راست برابر 10°
 و $\lambda = 0/3$. ۷۱
- ۶۰۳ نمودار چندک-چندک نرمال استاندارد برای باقی مانده‌های r_{Di}^* و r_{Mi}^*, r_{Ci}^*
 $n = 50, n = 100, n = 200$ ، درصد داده‌های سانسور از راست برابر 30°
 و $\lambda = 0/3$. ۷۲
- ۷۰۳ نمودار چندک-چندک نرمال استاندارد برای باقی مانده‌های r_{Di}^* و r_{Mi}^*, r_{Ci}^*
 $n = 50, n = 100, n = 200$ ، درصد داده‌های سانسور از راست برابر 0°
 $\lambda = 1/5$. ۷۳
- ۸۰۳ نمودار چندک-چندک نرمال استاندارد برای باقی مانده‌های r_{Di}^* و r_{Mi}^*, r_{Ci}^*
 $n = 50, n = 100, n = 200$ ، درصد داده‌های سانسور از راست برابر 10°
 و $\lambda = 1/5$. ۷۴

۹۰۳	نمودار چندک-چندک نرمال استاندارد برای باقی مانده‌های r_{Di}^* و r_{Mi}^*, r_{Ci}^*
	$n = 50, n = 100$ و $n = 200$ ، درصد داده‌های سانسور از راست برابر 30°
۷۵	و $\lambda = 1/5$
۱۰۴	نمودارهای $GD_i(\theta)$ و $LD_i(\theta)$ مربوط به داده‌های سرطان سینه
۸۴	
۲۰۴	نمودارهای مربوط به تأثیر موضعی کل و d_{max} برای طرح اغتشاش وزن-مورد
۸۵	
۳۰۴	نمودارهای مربوط به تأثیر موضعی کل و d_{max} برای طرح اغتشاش هم‌زمان
۸۵	متغیرهای پاسخ
۴۰۴	نمودارهای مربوط به تأثیر موضعی کل و d_{max} برای طرح اغتشاش متغیر
۸۶	توضیحی
۸۶	نمودار باقی مانده انحراف تعدیل یافته برای داده‌های سرطان سینه
۵۰۴	
۶۰۴	نمودار چندک-چندک نرمال استاندارد برای باقی مانده انحراف تعدیل یافته
۸۸	داده‌های سرطان سینه

فهرست جدول‌ها

۱۰۴	برآوردهای ماکسیمم درست‌نمایی برای پارامترهای مدل رگرسیون (۱۰۴)، برازش
۷۹	داده شده به داده‌های سرطان سینه
۲۰۴	برآوردهای جک‌نایف برای پارامترهای مدل رگرسیون (۱۰۴)، برازش داده شده
۷۹	به داده‌های سرطان سینه
۳۰۴	نتایج مربوط به روش بوت‌استرپ برای داده‌های سرطان سینه
۸۱	نتایج برآورد بیزی برای پارامترهای مدل
۵۰۴	نتایج برآورد بیزی برای پارامترهای مدل با استفاده از چگالی پیشین انقباضی
۸۱	ستیغی
۶۰۴	نتایج حاصل از شبیه‌سازی برای مقایسه چگالی ناآگاهی بخش و چگالی پیشین
۸۳	انقباضی ستیغی
۷۰۴	نتایج سنجش اثر داده‌های شماره ۳ و ۹۴ بر برآورد پارامترهای مدل رگرسیون
۸۷	وایبول نمایی شده لگاریتمی برای داده‌های سرطان سینه
۹۰	الف ۱. داده‌های واقعی مثال کاربردی فصل ۴

فهرست نمادها

صفحه	توضیح	نماد
۳.....	تابع درست‌نمایی	L
۸.....	لگاریتم تابع درست‌نمایی	ℓ
۲۹.....	توزیع وایبول نمایی‌شده لگاریتمی	LEW
۴۲.....	توزیع گاما	G
۴۲.....	توزیع معکوس گاما	IG
۴۷.....	مشتق مرتبه اول لگاریتم درست‌نمایی	$\dot{\ell}$
۴۷.....	مشتق مرتبه دوم لگاریتم درست‌نمایی	$\ddot{\ell}$

پیشگفتار

تحلیل بقاء از روش‌های پرکاربرد آماری است که توسط پژوهشگران بسیاری مورد مطالعه قرار می‌گیرد. یکی از روش‌های آماری در تحلیل بقاء، استفاده از مدل رگرسیون برای بررسی اثر متغیرهای توضیحی بر متغیر پاسخ طول عمر است که در بسیاری از پژوهش‌ها، این متغیر پاسخ به صورت سانسور شده بازه‌ای است.

لاولس^۱ در سال ۲۰۰۳ و کالت^۲ در سال ۲۰۰۳ بسیاری از روش‌های تحلیل بقاء در حالت پارامتری و ناپارامتری با انواع داده‌ها را مورد بررسی قرار داده‌اند. توزیع وایبول در روش‌های مختلف تحلیل بقاء کاربرد بسیاری دارد. تابع نرخ شکست این توزیع ثابت یا یکنوا است. مودهاکر^۳ و همکاران در سال ۱۹۹۵ توزیع وایبول نمایی شده را معرفی کرده‌اند. تابع نرخ شکست این توزیع علاوه بر حالت یکنوا، حالت‌های گودال-شکل و تک‌مدی را نیز شامل می‌شود، در نتیجه بسیاری از نمونه‌های واقعی را پوشش می‌دهد.

کوک^۴ در سال‌های ۱۹۷۷ و ۱۹۸۶ روش‌های تأثیر فراموضعی و تأثیر موضعی را برای تحلیل حساسیت مدل‌های رگرسیون پیشنهاد کرده است. اورتگا^۵ و همکاران در سال ۲۰۰۳ مسئله‌ی ارزیابی اثرات موضعی را در مدل رگرسیونی گامای تعمیم یافته با مشاهدات سانسور شده بررسی کرده‌اند. زی و وی^۶ در سال ۲۰۰۷ یک کاربرد تشخیص اثرات را در مدل رگرسیونی پواسون تعمیم یافته با داده‌های سانسور شده بر پایه‌ی روش حذف موردی و تحلیل اثرات موضعی توسعه داده‌اند. هاشیماتو^۷ و همکاران در سال ۲۰۱۰، توزیع وایبول نمایی شده لگاریتمی را برای تحلیل مدل‌های رگرسیونی با داده‌های سانسور شده بازه‌ای استفاده کرده‌اند.

در این پایان‌نامه، در فصل اول به بررسی بعضی از انواع سانسورها پرداخته و سپس

^۱Lawless

^۲Collet

^۳Mudholkar

^۴Cook

^۵Ortega

^۶Xie and Wei

^۷Hashimoto

پیش‌نیازهای موردنیاز در فصل بعد بیان می‌شود. در فصل دوم مدل‌های رگرسیون مورد استفاده برای داده‌های طول عمر تشریح می‌گردند. در فصل سوم به معرفی مدل رگرسیون وایبول نمایی شده لگاریتمی پرداخته می‌شود، برآورد پارامترهای آن، تحلیل حساسیت و تحلیل باقی‌مانده‌های مدل، مورد بررسی قرار می‌گیرد. در فصل چهارم یک مثال کاربردی از مدل رگرسیون معرفی شده مورد بررسی قرار می‌گیرد و نتایج گزارش می‌شود. برنامه‌های کامپیوتری این پایان‌نامه با نرم‌افزار R در پیوست آورده شده است.

