

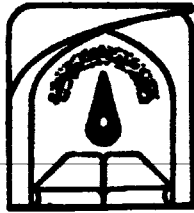
صلى الله عليه وسلم

۷۹۹۲
۶۵
۱۱/۲۵

۲۹۹۵۱

۱۳۸۱ / ۱ / ۲۰

کتابخانه تخصصی
پایه اول
شماره ثبت کتاب
۰۱۶۵۹۸



دانشگاه تربیت مدرس

دانشکده علوم پزشکی

بیان نامه کارشناسی ارشد آمار زیستی

کاربرد مدل رگرسیون دو جمله‌ای منفی در تعیین عوامل مؤثر بر حاملگی ناخواسته

نگارش:

شبنم کریمی

استاد راهنما:

دکتر انوشیروان کاظم نژاد

۳۹۹۶۱

استاد مشاور:

دکتر سیدمهدی سادات هاشمی

زمستان ۱۳۸۰

از اعضاء دانشمندان علم ایران
تسلیت داوران

«فرم تأییدیه اعضای هیأت داوران مندرج در پایان نامه کارشناسی ارشد»

بدینوسیله پایان نامه کارشناسی ارشد خانم شبنم کریمی

رشته: آمار حیاتی گرایش:

تقدیم می شود. اینجانبان نسخه نهائی این پایان نامه را از نظر فرم و محتوی بررسی و تأیید کرده و پذیرش آنرا برای تکمیل درجه کارشناسی ارشد پیشنهاد می کنیم.

نام و نام خانوادگی و امضاء اعضای هیأت داوران:

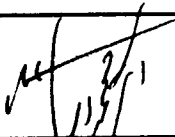
جناب آقای دکتر انوشیروان کاظم نژاد (ستاد راهنما)



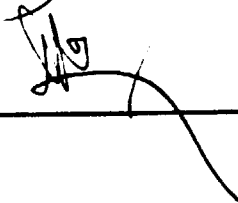
جناب آقای دکتر سید مهدی سادات هاشمی (استاد مشاور)



جناب آقای دکتر غلامرضا بابایی (نماینده تحصیلات تکمیلی)



جناب آقای دکتر محمود محمودی فراهانی (استاد ناظر)



جناب آقای دکتر سعید سقراط فقیهزاده (استاد ناظر)



بسمه تعالی

آیین‌نامه چاپ پایان‌نامه (رساله)‌های دانشجویان دانشگاه تربیت مدرس

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

- ماده ۱ در صورت اقدام به چاپ پایان‌نامه (رساله)ی خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثار علمی» دانشگاه اطلاع دهد.
- ماده ۲ در صفحه سوم کتاب (پس از برگ شناسنامه)، عبارت ذیل را چاپ کند:
«کتاب حاضر، حاصل پایان‌نامه کارشناسی ارشد/رساله دکتری نگارنده در رشته آمار زیستی است که در سال ۱۳۸۰ در دانشکده علوم پزشکی دانشگاه تربیت مدرس به راهنمایی سرکار خانم/جناب آقای دکتر کاظم نژاد، مشاوره سرکار خانم/جناب آقای دکتر سادات هاشمی 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 سوبت چاپ) را به «دفتر نشر آثار علمی» دانشگاه اهدا کند. دانشگاه می‌تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.
- ماده ۴ در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ‌شده را به عنوان خسارت به دانشگاه تربیت مدرس، تأدیه کند.
- ماده ۵ دانشجو تعهد و قبول می‌کند 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 عرضه‌شده نگارنده برای فروش، تأمین نماید.
- ماده ۶ اینجانب شبنم کریمی دانشجوی رشته آمار زیستی مقطع کارشناسی ارشد تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می‌شوم.

نام و نام خانوادگی: شبنم کریمی

تاریخ و امضا: ۲۵ / ۱۰ / ۸۰

ای ذکر جانفزای تو مفتاح بابها وی نام دلفروز تو زیب کتابها

سپاس خدای را که منت بندگی بر من نهاد و توفیق گذراندن این دوره از زندگی را به من عطا نمود.

تقدیم به پدر و مادر مهربانم، که روشنی فکرشان چراغ راهنمای من به سوی موفقیتها شد و فروغ چشمانشان، روشنی بخش راه زندگی ام.

تقدیم به پدر و مادر بزرگوار همسرم، که همواره مرا همچون فرزندان خویش مورد لطف قرار دادند.

تقدیم به خواهر و برادر خوبم، که همواره آرزوی موفقیت روزافزونشان را دارم.

و

تقدیم به همسر عزیزم، که با یاری، شکیبایی و همدلی موجب آسودگی خاطر و ثبات قدمم گردید.

باتشکر فراوان از:

● استاد بزرگوار و ارجمند

جناب آقای دکتر انوشیروان کاظم نژاد که به عنوان استاد راهنما در تمامی مراحل این تحقیق مرا یاری نمودند.

● جناب آقای دکتر سید مهدی سادات هاشمی که در امر مشاوره پایان نامه از نظرات و مشاوره ایشان بهره مند شدم.

● اساتید محترم گروه آمار حیاتی جناب آقای دکتر غلامرضا بابایی و جناب آقای دکتر سقراط فقیه زاده که طی دوره تحصیل همواره مرا راهنما بودند.

● جناب آقای دکتر محمود محمودی که قبول زحمت نموده و داوری این پایان نامه را به عهده گرفتند.

چکیده

موضوع تحقیق حاضر « کاربرد مدل رگرسیون دو جمله‌ای در تعیین عوامل مؤثر بر

حاملگی ناخواسته » می باشد و متغیر وابسته تعداد حاملگی‌های ناخواسته در نظر گرفته شده است.

در آنالیز رگرسیون در حالتی که متغیر وابسته گسسته و نامنفی باشد، مدل رگرسیون پواسن مورد استفاده قرار می گیرد. مدل رگرسیون پواسن به عنوان نمونه‌ای از مدل‌های تعمیم یافته خطی است که مدل‌های رگرسیونی را به خانواده نمایی توزیع‌هایی بسط می دهد که هر دو توزیع نرمال و پواسن را شامل می شوند. شرط اصلی استفاده از مدل رگرسیون پواسن معادل بودن میانگین و واریانس متغیر وابسته می باشد. وقتی که میانگین و واریانس داده ها به طور تقریبی برابر نباشند، مدل پواسن برآوردهای ناصحیحی از واریانس جملات و استنباط‌هایی گمراه کننده درباره رگرسیون ایجاد می کند. برای حل این مشکل می توان از مدل رگرسیونی دو جمله‌ای منفی استفاده کرد.

به منظور تعیین مهمترین عوامل خطر ساز در حاملگی ناخواسته و تعیین برآورد پارامترها از مدل‌های رگرسیونی پواسن و دو جمله‌ای منفی استفاده شده و مقایسه ای بین آنها صورت گرفته است. از بین کلیه متغیرهای مورد بررسی، متغیرهای سطح تحصیلات مادر، تعداد فرزندان زنده دختر، تعداد فرزندان زنده پسر و استفاده از روش پیشگیری مؤثر بودند و متغیرهای سن، سطح تحصیلات پدر و آگاهی قبلی از روشهای تنظیم خانواده از نظر آماری از مدل حذف شدند و اثرات متقابل نیز وارد مدل نشدند.

واژه‌های کلیدی: رگرسیون دو جمله‌ای منفی - رگرسیون پواسن - حاملگی ناخواسته .

فصل اول: مقدمه و بیان مسأله

۲	۱-۱ مقدمه
۴	۲-۱ تعریف مسئله و بیان اصلی تحقیق
۵	۳-۱ فرضیه های تحقیق
۵	۴-۱ بیان اهداف تحقیق
۵	۵-۱ اهمیت پژوهش
۸	۶-۱ نوع مطالعه
۸	۷-۱ جامعه آماری
۸	۸-۱ حجم اطلاعات مورد بررسی
۹	۹-۱ واحد آماری
۹	۱۰-۱ چگونگی دستیابی به اطلاعات
۹	۱۱-۱ تعریف متغیرها
۹	۱-۱۱-۱ متغیر وابسته
۱۰	۲-۱۱-۱ متغیرهای مستقل

فصل دوم: حاملگی ناخواسته

۱۳	۱-۲ مقدمه
۱۵	۲-۲ حاملگی ناخواسته
۱۷	۳-۲ پیامدهای حاملگی ناخواسته
۱۷	۱-۳-۲ احساس ناخوشایند مادر نسبت به حاملگی
۱۸	۲-۳-۲ سقط جنین
۱۸	۳-۳-۲ کودک ناخواسته
۲۰	۴-۳-۲ غفلت از کودک
۲۲	۱-۴-۳-۲ عدم تغذیه انحصاری با شیر مادر
۲۲	۲-۴-۳-۲ عدم وزن گیری و رشد کافی
۲۳	۳-۴-۳-۲ عدم مراجعه منظم جهت انجام واکسیناسیون
۲۴	۵-۳-۲ افزایش مرگ و میر کودکان
۲۵	۶-۳-۲ پیامدهای رفتاری
۲۶	۴-۲ مروری بر مطالعات انجام شده

فصل سوم : خانواده توزیع های نمایی

۳۴	۱-۳ مقدمه
۳۵	۲-۳ خانواده نمایی
۳۷	۳-۳ تعریف بسندگی
۳۸	۴-۳ آماره های بسنده (کافی)
۳۹	۵-۳ توزیع پواسن
۴۴	۶-۳ توزیع دو جمله ای منفی
۴۵	۱-۶-۳ نمادهای مختلف توزیع دو جمله ای منفی
۴۸	۲-۶-۳ تعریف رسمی توزیع دو جمله ای منفی
۵۲	۳-۶-۳ توزیع دو جمله ای منفی به عنوان یک توزیع آمیخته
۵۲	۱-۳-۶-۳ تعریف توزیع آمیخته
۵۳	۲-۳-۶-۳ توسعه مفهوم آمیختگی

فصل چهارم : تحلیل داده های شمارشی و نرخها

۵۶	۱-۴ مقدمه
۵۹	۲-۴ الگوهای تعمیم یافته خطی
۶۱	۱-۲-۴ تعریف مدل خطی
۶۳	۳-۴ آنالیز رگرسیون شمارش ها و نرخ ها
۷۲	۱-۳-۴ رگرسیون پواسن
۷۸	۱-۱-۳-۴ رگرسیون پواسن بیش پراکنده
۸۷	۲-۳-۴ رگرسیون دو جمله ای منفی

فصل پنجم : روشهای مدل سازی رگرسیون پواسن و دو جمله ای منفی

۹۷	۱-۵ مدل رگرسیون پواسن
۹۹	۱-۱-۵ برآورد حداکثر درست نمایی پارامترها در مدل پواسن
۱۰۳	۲-۵ مدل رگرسیون دو جمله ای منفی
۱۰۶	۱-۲-۵ برآورد حداکثر درست نمایی پارامترها در مدل دو جمله ای منفی
۱۰۹	۳-۵ آزمون های آماری رگرسیون پواسن و رگرسیون دو جمله ای منفی
۱۰۹	۱-۳-۵ آزمون χ^2
۱۱۰	۲-۳-۵ آزمون معنی داری ضرایب

۱۱۱	۳-۳-۵ آزمون نسبت درستی مدلی پواسن به مدل دو جمله ای منفی
۱۱۱	۴-۳-۵ مقدار α
۱۱۲	۴-۵ تفسیر ضرایب مدل رگرسیون
۱۱۵	۵-۵ مقایسه مدل های کاهش و استنباط درباره ضرایب آنها
۱۱۵	۶-۵ روش مدل سازی
۱۱۶	۷-۵ روش گام به گام (پیشرو - پسرو)

فصل ششم: نتایج، بحث و پیشنهادات

۱۲۱	۱-۶ نحوه برآزش مدل
۱۲۱	۱-۱-۶ مدل رگرسیون پواسن
۱۲۴	۲-۱-۶ مدل رگرسیون دو جمله ای منفی
۱۲۷	۲-۶ برآورد خطر نسبی متغیرهای موجود در مدل
۱۲۲	۳-۶ نتایج
۱۳۳	۴-۶ پیشنهادات
۱۳۵	۵-۶ بحث
۱۳۹	فهرست منابع
۱۳۳	ضمیمه الف (رگرسیون کرنل (kernel Regression))

فصل اول

مقدمه و بیان مسأله

۱-۱ مقدمه

بخش وسیعی از آمار، زیست‌سنجی و اقتصاد، به تحلیل داده‌های پاسخ گسسته مربوط می‌شود. در بررسی‌های آماری در مواردی که متغیرهای مستقل وابسته از نوع پیوسته باشند از روش‌های رگرسیونی استفاده می‌شود. در مواردی که متغیرهای مستقل گسسته و متغیرهای پاسخ پیوسته باشند از روش آنالیز واریانس و اگر متغیرهای مستقل برخی گسسته و برخی پیوسته و متغیر پاسخ از نوع پیوسته باشد از آنالیز کوواریانس استفاده می‌شود. ولی در مواردی که متغیر پاسخ گسسته باشد (متغیرهای مستقل از هر نوع که باشند) مدل‌های تعمیم یافته خطی را به کار می‌بریم. اینگونه موارد بخش عمده‌ای از بررسی‌های آماری به ویژه تحلیل داده‌های زیستی را شامل می‌شود.

مدل رگرسیون پواسن و دو جمله‌ای منفی نمونه‌هایی از مدل‌های تعمیم یافته خطی اند که برای آنالیز داده‌های شمارشی مورد استفاده قرار می‌گیرند.

چون یکی از اهداف این تحقیق تعیین عوامل مؤثر بر حاملگی ناخواسته می‌باشد لذا با استفاده از مدل‌های رگرسیونی پواسن و دو جمله‌ای منفی به تجزیه و تحلیل داده‌های مربوط به حاملگی ناخواسته می‌پردازیم.

یکی از اهداف آمار حیاتی یا پزشکی اجتماعی شناسایی مشکلات و نیازهای بهداشتی

جامعه و ارائه روش های قابل اجرا برای حل آن می باشد و هدف دیگر آن، استقرار سلامتی در جامعه است.

با وجود پیشرفت های انجام شده در زمینه علم بارداری و زایمان هنوز هم مواردی وجود دارد که به طور گسترده سلامت و تندرستی مادران را به خطر می اندازد. مسمومیت های حاملگی، خونریزی ها و عفونت ها به دلایل مختلف از جمله سقط جنین به دنبال وقوع حاملگی ناخواسته و برنامه ریزی نشده، بارداری های مکرر و با فاصله کوتاه که اکثراً در زمان دلخواه زوجین رخ نمی دهند از جمله مواردی هستند که در این رابطه مطرح می باشند. عدم جلوگیری مؤثر از زایمان های مکرر و بدون برنامه ریزی، بدون تردید علاوه بر مادران، میزان مرگ و میر و آسیب های وارده بر کودکان را نیز به طور چشم گیری افزایش می دهد. بیش از ۳۰۰ میلیون زن در سراسر جهان که در سنین باروری قرار دارند (حدود $\frac{1}{3}$ کل زنان بارور جهان) نمی خواهند حاملگی مجددی داشته باشند یا اینکه حاملگی فعلی آنها به طور ناخواسته اتفاق افتاده است.

با توجه به عوارض گسترده حاصل از وقوع حاملگی های ناخواسته، چنانچه به بررسی دقیق عوامل مؤثر بر این قبیل بارداری ها پرداخته شود، ارائه پیشنهادات در زمینه راه های پیشگیری از این نوع بارداری ها جهت قشر عظیمی از زنان در معرض خطر وقوع بارداری ناخواسته نیز بسیار دقیق تر و جامع تر خواهد بود.

در این تحقیق بر آنیم که با کمک مدل های آماری مهمترین عوامل مؤثر بر حاملگی ناخواسته را شناسایی کنیم.

۲-۱ تعریف مسأله و بیان اصلی تحقیق

در آنالیز رگرسیون در حالتی که متغیر وابسته شمارشی و نامنفی باشد، مدل رگرسیون پواسن مورد استفاده قرار می گیرد. یعنی وقتی که متغیر مورد نظر گسسته، نامنفی و اسپورادیک باشد.

محدوده اصلی مدل رگرسیون پواسن بر این شرط استوار است که واریانس متغیر وابسته با میانگین آن برابر باشد، اما در بسیاری از پدیده ها داده ها بسیار پراکنده اند. به عبارت دیگر ممکن است واریانس به طور معنی داری بزرگتر از میانگین باشد. وقتی که میانگین و واریانس داده ها به طور تقریبی مساوی نباشند، واریانسهای ضرایب برآورده شده در مدل پواسن اریب خواهند بود. در این حالت دیگر برازش رگرسیون پواسن بر روی داده ها مناسب نمی باشد ولی می توان این محدودیت را با استفاده از توزیع دو جمله ای منفی از میان برداشت.

توزیع دو جمله ای منفی برای داده های گسسته و نامنفی مناسب بوده و به برابری میانگین و واریانس حساس نیست. پس به لحاظ مشابه بودن فرم فضای نمونه ای، می توان از توزیع دو جمله ای منفی به جای توزیع پواسن استفاده کرد و برآوردهای دقیقتری را به دست آورد. تعداد حاملگی های ناخواسته در یک دوره زمانی مشخص یک متغیر تصادفی گسسته می باشد و ماهیت آن به صورت احتمالی است و مقدار صحیح نامنفی می باشد. پس مدل نمونه گیری ما مدل پواسن است زیرا چنین توزیعی زمانی رخ می دهد که مشاهدات در طول دوره ای از زمان، بدون آگاهی پیشین از تعداد کل مشاهدات به عمل آید. به منظور مدل بندی، بررسی و تعیین عوامل مؤثر بر حاملگی ناخواسته از مدل رگرسیون پواسن و مدل رگرسیون دو جمله ای منفی استفاده می شود. در این تحقیق متغیر پاسخ یعنی تعداد حاملگی های ناخواسته متغیری

شمارشی است با توزیع پواسن و تأثیر متغیرهای دموگرافیک و بهداشتی بر وقوع حاملگی های ناخواسته یا استفاده از مدل های رگرسیونی پواسن و دو جمله ای منفی مورد بررسی قرار می گیرد.

۳-۱ فرضیه های تحقیق

الف: مدل رگرسیون دو جمله ای منفی برازش بهتری رانسبت به رگرسیون پواسن برای مدل بندی عوامل مؤثر بر حاملگی ناخواسته می دهد.

ب: عوامل دموگرافیک و بهداشتی در وقوع حاملگی های ناخواسته مؤثر هستند.

۴-۱ بیان اهداف تحقیق

الف: هدف کلی: عبارت است از کاربرد مدل رگرسیون دو جمله ای منفی در تعیین عوامل مؤثر بر حاملگی ناخواسته.

ب: اهداف جزئی:

۱- به دست آوردن مدل رگرسیونی مناسبتر نسبت به مدل رگرسیون پواسن برای به دست آوردن برآوردهایی دقیقتر.

۲- تعیین عوامل مؤثر بر حاملگی ناخواسته.

۵-۱ اهمیت پژوهش

ساده ترین مدل رگرسیونی که برای تجزیه و تحلیل داده های شمارشی مورد استفاده قرار