



Comparison between Efficiency of Poisson Regression Model and Negative Binomial Regression in the Analysis of Factors Affecting Mortality from Cardiovascular Diseases in Yazd Province in 2017

Mohammad Mirjani Arjenan(MS.c.)¹, Mohsen Askarshahi(Ph.D.)², Mahmud Vakili(Ph.D.)³

1.M.Sc. Student in Biological Statistics, International Campus, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

2. Corresponding Author :Associate Professor, Department of Statistics and Epidemiology, School of Public Health., Shahid Sadoughi University of Medical Sciences, Yazd, Iran. E-mail: moasbio@gmail.com Tel:09125799189

3. Associate Professor, Department of Health Monitoring, School of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

Abstract

Introduction: Despite the advances in cardiovascular diseases, death caused by these diseases is still considered as the leading cause of mortality. In this study, some of the effective factors on the deaths caused by cardiovascular diseases were investigated

Methods: This cross-sectional analytical study investigated the efficacy of Poisson regression models and negative binomial regression models on factors affecting mortality from cardiovascular diseases. The death data were extracted from the death registration system for Yazd province in 2017. Gender, age, education, occupation, location, and city of death were also extracted for each deceased. The two regression models were then fitted to the data

Results: A total of 5,015 deaths were recorded, of which 1,642 were due to cardiovascular diseases. Cardiovascular disease mortality rates were significant using negative binomial regression in terms of the educational variables, place of residence, type of residence, and age. Death rates caused by cardiovascular diseases were not significant for age and occupational, educational, and residential variables.

Conclusion: If the time of death is considered as an offset variable, the regression model of two negative sentences is more effective in showing the factors affecting death due to cardiovascular diseases according to AIC(Akaike Information Critcrion) and BIC(Bayesian Information Critcrion) criteria. In the case that the total number of deaths is considered as the offset variable, the Poisson regression model is more efficient.

Keywords: Cardiovascular Disease, Negative Binomial Regression, Poisson Regression, Performance, Akaike Score, Bayesian Score

Conflict of interest: The authors declared that there is no conflict of interest.



This Paper Should be Cited as:

Author: Mohammad Mirjani Arjenan, Mohsen Askarshahi, Mahmud Vakili.
Comparison between Efficiency of Poisson RegressionToloobehdasht
Journal.2021;19(5):55-67.[Persian]



مقایسه کارآیی مدل رگرسیون پواسون و رگرسیون دو جمله ای منفی در تحلیل عوامل موثر بر مرگ و میر ناشی از بیماریهای قلبی و عروقی در استان یزد در سال ۱۳۹۶

نویسندگان: محمد میرجانی ارجنان^۱، محسن عسکری شاهی^۲، محمود وکیلی^۳

۱. دانشجوی کارشناسی ارشد آمار زیستی، پردیس بین الملل، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی، یزد، ایران.

۲. نویسنده مسئول: دانشیار گروه آمار و اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی، یزد، ایران. تلفن تماس: ۰۹۱۲۵۷۹۹۱۸۹

Email: moasbio@gmail.com

۳. دانشیار گروه پزشکی اجتماعی، دانشکده پزشکی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی، یزد، ایران.

طلوع بهداشت

دو ماهنامه علمی پژوهشی
دانشکده بهداشت یزد
سال نوزدهم
شماره پنجم
آذر و دی ۱۳۹۹
شماره مسلسل: ۸۳

تاریخ وصول: ۱۳۹۸/۱۱/۲۶

تاریخ پذیرش: ۱۳۹۹/۰۲/۲۸

چکیده

مقدمه: با وجود پیشرفت‌های انجام شده در زمینه بیماری‌های قلبی و عروقی هنوز هم مرگ ناشی از این بیماری در صدر علت‌های مرگ می‌باشد. در این مطالعه به بررسی دقیق برخی از این عوامل موثر بر مرگ ناشی از بیماری‌های قلبی و عروقی پرداخته می‌شود.

روش بررسی: مطالعه‌ای تحلیلی از نوع مقطعی که در آن کارآیی مدل‌های رگرسیون پواسون و دو جمله ای منفی در ارتباط با عوامل موثر بر مرگ و میر ناشی از بیماری‌های قلبی و عروقی بررسی می‌شود. اطلاعات مرگ از سامانه ثبت مرگ برای سال ۱۳۹۶ استان یزد استخراج شد. جنس، سن، تحصیلات، شغل، موقعیت مکانی، شهرستان محل فوت نیز برای هر کدام از متوفیان استخراج شد. سپس این دو مدل رگرسیونی به داده‌ها برازش داده شد.

یافته‌ها: تعداد ۵۰۱۵ مرگ ثبت شده است که از این تعداد ۱۶۴۲ فوت در اثر بیماری‌های قلبی و عروقی اتفاق افتاده بود. نرخ مرگ در اثر بیماری قلبی و عروقی با استفاده از رگرسیون دو جمله ای منفی برای متغیرهای تحصیلات، محل سکونت، نوع سکونت و سن معنی دار شد. نسبت مرگ به علت بیماری قلبی و عروقی برای متغیر سن معنی دار و برای متغیرهای شغل، تحصیلات، نحوه سکونت و محل سکونت معنی دار نشد.

نتیجه گیری: در صورتیکه زمان وقوع مرگ به عنوان متغیر offset در نظر گرفته شود، مدل رگرسیون دو جمله ای منفی برای نشان دادن عوامل موثر بر مرگ به علت بیماری قلبی و عروقی با توجه به معیار اطلاع آکائیک (Akaike Information Criterion) و معیار اطلاع بیزی (Bayesian Information Criterion) کارتر می‌باشد و در صورتیکه کل افراد فوت شده به عنوان متغیر offset در نظر گرفته شود، مدل رگرسیون پواسون کارتر می‌باشد.

واژه‌های کلیدی: مرگ ناشی از بیماری‌های قلبی و عروقی، رگرسیون دو جمله ای منفی، رگرسیون پواسون

این مقاله برگرفته از پایان نامه کارشناسی ارشد پردیس بین الملل، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد می‌باشد.



مقدمه

بیماری قلبی و عروقی هم چنان از شایع ترین علت مرگ در بیشتر کشورهای جهان محسوب می شود (۱). بیماری های قلبی و عروقی در اغلب کشورهای جهان از جمله کشور ایران به عنوان یکی از مهمترین عوامل تهدید کننده سلامت انسان ها شناخته می شود (۲). این بیماری همچنین اصلی ترین عامل مرگ و میر در ایران محسوب می شود. بیماری های قلبی عروقی یکی از بیماری های غیرواگیر قابل پیشگیری انسان به شمار می آیند (۳،۴).

سازمان بهداشت جهانی برآورد کرده است که اگر روند فعلی ادامه پیدا کند و اقدامی در جهت بهبود بیماران قلبی و عروقی صورت نگیرد تا سال ۲۰۲۰ در سراسر جهان ۲۵ درصد از سال های سلامت زندگی افراد به علت بیماری های قلبی عروقی از دست می رود (۵،۶). شیوع بالای بیماری قلبی و عروقی، اهمیت مقابله با این بیماری را آشکار می سازد و به همین دلیل شناسایی و بررسی عوامل خطر ساز زیستی و روان شناختی و پیشگیری و درمان در این حیطه ضروری به نظر می رسد (۷). آنالیز رگرسیون چندگانه یکی از روش های آماری می باشد که برای برآورد عوامل خطر ساز و سهم اثرات آن عوامل به طور چشمگیری استفاده می شود (۸). در آنالیز رگرسیون در حالتی که متغیر وابسته گسسته و نامنفی باشد، مدل رگرسیون پواسن مورد استفاده قرار می گیرد (۹). محدوده اصلی مدل رگرسیون پواسن بر این شرط استوار است که واریانس متغیر وابسته با میانگین آن برابر باشد (۱۰). مدل های داده های شمارشی به طور گسترده در مطالعات تجربی استفاده می شود و مدل های رایج برای این داده ها مدل پواسون، مدل دو جمله ای و دو جمله ای منفی

می باشد (۱۱،۱۲). اما در بسیاری از پدیده ها داده ها بسیار پراکنده اند، به عبارت دیگر ممکن است واریانس به طور معنی داری از میانگین بزرگتر باشد. وقتی که واریانس و میانگین متغیر پاسخ با هم برابر نباشند، واریانس های ضرایب برآورد شده در مدل پواسن اریب خواهند بود که در این حالت برازش رگرسیون پواسن بر روی داده ها مناسب نمی باشد. مدل دو جمله ای منفی به برابری واریانس و میانگین جامعه حساس نمی باشد، پس با توجه به تشابه فرم فضای نمونه ای می توان از مدل دو جمله ای منفی به جای مدل پواسن استفاده کرد (۱۳). Ardiles و همکاران در سال ۲۰۱۸ مدل رگرسیون دو جمله ای منفی را برای داده های بیماران بستری استفاده کردند. آن ها تعداد بیماران تنفسی و خونی را با توجه به سن با آلاینده های هوا بررسی کردند. بهترین مدل برای توصیف این داده ها رگرسیون دو جمله ای منفی عنوان شد و آلاینده CO جهت بیماری خونی در سالمندان و بیماری تنفسی جهت افراد ۱۹-۱۰ سال عامل خطر معرفی شد در حالیکه عامل خطر برای جمعیت ۶۴-۲۰ سال NO2 معرفی شد (۱۴). Waenerlund و همکاران در سال ۲۰۱۸ در سوئد ارتباط بین سطح درآمد و تحصیلات را بین زنان و مردان ۳۸ تا ۶۲ ساله بیماران بستری بیمارستان های وابسته به بیماری قلبی و عروقی سنجیدند. سطح درآمد و تحصیلات در ۶ زیرگروه برای زنان و مردان بررسی شد. بیماری های قلبی و عروقی برای زنان و مردان برای نابرابری های اقتصادی و آموزشی، پایدار بود ولی برای مردان کاهش داشت. نتایج افزایش بیماری های قلبی و عروقی را در شرایط نابرابری های اجتماعی نشان داد (۱۵). Correia و همکاران در سال ۲۰۱۸ با استفاده از توزیع دو جمله ای منفی به عوامل موثر



تصادفاتی که در بخشی از بزرگراهی در تایوان در یک دوره زمانی یکساله اتفاق افتاده است را مدل بندی کرده است. نتایج نشان داد شبکه عصبی مصنوعی برای پیش بینی یک یا بیشتر از یک تصادف عملکرد بهتری دارد درحالی که اگر انباشتگی در مقدار صفر باشد رگرسیون دو جمله ای منفی با ۹۷٪ پیش بینی درست شبکه عصبی مصنوعی با ۹۰٪ پیش بینی عملکرد بهتری دارد (۱۸). Wong و Famoy در سال ۲۰۰۴ مدل رگرسیون پواسن تعمیم یافته را چنان گسترش دادند که متغیر پاسخی را پیش بینی کند که تحت تاثیر یک یا چند متغیر توضیحی است. مدل رگرسیون پواسن تعمیم یافته برای مدل سازی داده های شمارشی مناسب است که بیش پراکندگی یا کم پراکندگی را نشان می دهد. پارامترهای رگرسیون توسط ماکزیمم درست نمایی برآورد شد. نتایج مطالعه نشان داد که اثر سانسور روی خطای معیار پارامترها با استفاده از هموارسازی انجام می شود.

غلامرضا شریفی راد و همکاران در سال ۱۳۸۵ ارتباط بیماری های قلبی و عروقی در سالمندان را با سابقه فعالیت جسمانی بررسی کردند نتایج این مطالعه نشان داد که شیوع بیماری های قلبی عروقی و فشار خون بالا در سالمندان عضو کانون بازنشستگان اصفهان بیش از حد انتظار است. هم چنین این مطالعه مشخص کرد که میزان پیاده روی، تنیس روی میز، نرمش، دوچرخه سواری و کوه پیمایی در میانسالی در کاهش به بیماری های قلبی عروقی مؤثر است. مطالعه نشان داد که میزان پیاده روی، نرمش و دویدن در میانسالی بین دو جنس تقریباً یکسان بوده و میزان فعالیت جسمانی به طور مستقیم با سطح تحصیلات ارتباط دارد. به عبارتی هر چه سطح تحصیلات بالاتر باشد میزان فعالیت جسمانی هم بیشتر است. اما میزان پیاده روی،

برای تعداد بیماران مراجعه کننده به واحد اورژانس سرپایی شهر ژنو پرداختند. این داده ها از ماه مه ۲۰۱۳ تا اوت ۲۰۱۷ جمع آوری شد و تعداد پذیرش بیماران در دوره هایی که مسابقه تنیس برگزار می شدند با دوره هایی که مسابقات تنیس برگزار نمی شدند بررسی شد. کاهش معنی داری در تعداد پذیرش در زمان هایی که مسابقه تنیس برگزار می شد در مقابل زمان هایی که مسابقه ای نبود، مشاهده شد. این کاهش در مسابقات نیمه نهایی و فینال هم بیشتر مشاهده شد.

هم چنین کاهش تعداد پذیرش بیماران برای سنین ۶۴ تا ۲۶ ساله نیز مشهود بود و تغییری در نرخ پذیرش برای ساعت های قبل و بعد از پذیرش مشاهده نشد (۱۶). مقایسه ی توزیع پواسن تعمیم یافته و توزیع دو جمله ای منفی در سال ۲۰۰۵ توسط آقای Joe و همکارانش انجام شد. آن ها به این نتیجه رسیدند که برازش داده های شمارشی مدل پواسن تعمیم یافته و دو جمله ای منفی اغلب شبیه هم می باشد و برای تشخیص تفاوت های آن ها، تابع جرم احتمال و اریبی مدل پواسن تعمیم یافته و دو جمله ای منفی را با یکی از دو گشتاور ثابت مقایسه می کنیم. آن ها در بعضی موارد تفاوت های کمی دارند. با استفاده از داده های شمارشی با کسر زیادی از صفر بین توزیع پواسن تعمیم یافته و دو جمله ای منفی تفاوت وجود دارد. این مقایسه احتمالی به انتخاب یک توزیع برازش شده ی بهتر برای مدل سازی داده های شمارشی با دنباله ای از راست کشیده شده کمک می کند (۱۷).

در مطالعه ای که Li-yen chang در سال ۲۰۰۵ تحت عنوان تجزیه و تحلیل فراوانی تصادفات آزاد راه: رگرسیون دو جمله ای منفی در مقابل شبکه عصبی مصنوعی انجام داده است تعداد



MASS کدهای مربوط به محاسبات را نوشتیم. سپس مدل های رگرسیون پواسون و رگرسیون دو جمله ای منفی به داده ها برازش داده شدند. برای بر آورد پارامترهای مدل از روش درست نمایی ماکزیمم استفاده کردیم. هم چنین برای نشان دادن مدل مناسب تر از آماره لگاریتم و معیار اطلاع آکائیک (AIC) و (BIC) استفاده کردیم.

جهت انجام آنالیز و تحلیل داده ها پس از تکمیل اطلاعات مورد نیاز و تعریف متغیرهای مطالعه و کدگذاری آن ها، داده ها وارد نرم افزار R شدند و از پکیج های VGAM، MASS، ggplot2 و countreg استفاده شد. هم چنین سطح معنی داری ۰/۰۵ در نظر گرفته شد. این مطالعه پس از بررسی در کمیته اخلاق دانشکده بهداشت دانشگاه علوم پزشکی شهید صدوقی یزد با اخذ کد IR.SSU.SPH.REC.1398.130 مورد تایید قرار گرفت.

یافته ها

هدف از انجام این مطالعه مقایسه کارآیی مدل رگرسیون پواسون و دو جمله ای منفی برای تعیین عوامل موثر بر مرگ و میر ناشی از بیماری های قلبی و عروقی استان یزد در سال ۱۳۹۶ است. در این مطالعه تعداد کل مرگ و میرها در سال ۱۳۹۶ در استان یزد ۵۰۱۵ مورد بود که در این میان تعداد ۱۶۴۲ مورد مرگ بر اثر بیماری قلبی و عروقی داشتیم که ۴۹/۳٪ (۸۰۹ نفر) از این فوت شدگان مرد ۵۰/۷٪ زن می باشند. هم چنین از این تعداد ۵۵٪ ساکن یزد و بقیه ساکن سایر شهرستان های استان بوده اند. ۷۹/۷٪ از افرادی که به علت بیماری قلبی و عروقی فوت کرده اند در مناطق شهری و ۲۰/۳٪ در مناطق روستایی ساکن بوده اند. ۴۶/۵٪ از افرادی که شغل آن ها در گواهی فوت

دوچرخه سواری، کوه پیمایی و دویدن در میانسالی در تمام سطوح تحصیلی تقریباً یکسان صورت گرفته است (Dogra ۱۹). و همکاران در یک مطالعه نشان دادند بین بیماری های قلبی و عروقی و آسم در بزرگسالان ارتباط معنی داری وجود دارد. در این مطالعه با استفاده از ۷۴۳۴۲ شرکت کننده با رنج سنی ۴۵ تا ۷۰ سال و با استفاده از رگرسیون لجستیک چندگانه نشان داده شد که احتمال ابتلا به بیماری قلبی برای فرد آسمی ۴۳ درصد بیشتر از یک فرد غیر آسمی است. هم چنین هیچ ارتباط آماری بین سن شروع فشار خون و بیماری قلبی مشاهده نشد (۲۰).

روش بررسی

مطالعه حاضر مطالعه ای تحلیلی از نوع مقطعی است که به منظور کارآیی مدل رگرسیون پواسون و دو جمله ای منفی در ارتباط با عوامل موثر بر مرگ و میر ناشی از بیماری های قلبی و عروقی استان یزد در سال ۱۳۹۶ می باشد

کلیه ی افراد فوت شده به علت بیماری های قلبی و عروقی و سایر علت ها در سال ۱۳۹۶ استان یزد در مطالعه وارد شدند. مطالعه حاضر به منظور مقایسه کارآیی مدل رگرسیون پواسون و دو جمله ای منفی در ارتباط با عوامل موثر بر مرگ و میر ناشی از بیماری های قلبی و عروقی در استان یزد در سال ۱۳۹۶ می باشد. در این مطالعه کلیه افرادی که در سال ۹۶ به علت بیماری قلبی و عروقی و سایر بیماری ها فوت کرده اند بر اساس گواهی فوت صادر شده وارد مطالعه شدند.

پس از جمع اوری داده ها از سامانه ثبت مرگ در قالب excel2012 و تبدیل آن ها به داده های شمارشی با استفاده از Spss نسخه ۲۰ به منظور محاسبات آماری، داده ها در نرم افزار R فراخوانی شدند و با استفاده از پکیج نرم افزاری



می دهد. ابتدا تعداد مرگ و میر های گزارش شده بر اثر بیماری قلبی و عروقی در سال ۹۶ را به تفکیک متغیر های ماه، شغل، سطح تحصیلات، نحوه سکونت (شهری یا روستایی)، محل سکونت و سن را محاسبه کردیم. سپس تعداد کل موارد مرگ و میر و تعداد مرگ و میر در اثر بیماری قلبی و عروقی در هر زیر گروه را شمرده ایم.

سن	مرد	زن	ماه	مرد	زن	شغل	مرد	زن	محل سکونت	مرد	زن	وضعیت سکونت	مرد	زن
<۵۵	۹	۱۴	فروردین	۱۳	۸	شاغل	۱۵	۳۸	محل سکونت	۵۵	۸۴	روستایی	۲۰	۲
	۷													
≥۵۵	۹۱	۱۴	اردیبهشت	۱۶	۱۲	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۱۳	۱۳	خرداد	۱۸	۱۲	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۸	۷	تیر	۱۴	۱۲	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۱۶	۱۲	مرداد	۱۶	۱۲	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۷	۵	شهریور	۱۳	۱۲	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۱۸	۱۲	مهر	۱۲	۱۲	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۷	۸	آبان	۱۶	۱۲	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۱۰	۱۶	آذر	۱۰	۱۶	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۴													
	۱۷	۱۷	دی	۱۷	۱۷	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۱۰	۵	بهمن	۱۲	۱۳	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۱۲	۱۳	اسفند	۱۲	۱۳	شاغل	۴۶	۲	محل سکونت	۴۵	۶۹	روستایی	۳۱	۳
	۸	۴												

ثبت شده و به علت بیماری های قلبی و عروقی مرده اند، شاغل و ۵۳/۵٪ خانه دار ثبت شده است. سن ۹۱٪ از افرادی که به علت بیماری قلبی و عروقی فوت کرده اند بیشتر و بالاتر از ۵۵ سال و ۹٪ پایین تر از ۵۵ سال بوده است. جدول ۱ توزیع فراوانی مرگ و میر در اثر بیماری قلبی و عروقی بر حسب متغیرهای را نشان

جدول ۱: فراوانی مرگ و میر بر اثر بیماریهای قلبی و عروقی به تفکیک متغیرها در سال ۱۳۹۶ در استان یزد

جنسیت	مرد	زن	شغل	محل سکونت	وضعیت سکونت
فرا	۸۰	۸۳	شاغل	۸۴	روستایی
وا	۹	۳	شاغل	۸۴	روستایی
نی	۳	۳۸	شاغل	۸۴	روستایی
فرا	۸۰	۸۳	شاغل	۸۴	روستایی
وا	۹	۳	شاغل	۸۴	روستایی
نی	۳	۳۸	شاغل	۸۴	روستایی
جنسیت	مرد	زن	شغل	محل سکونت	وضعیت سکونت
فرا	۸۰	۸۳	شاغل	۸۴	روستایی
وا	۹	۳	شاغل	۸۴	روستایی
نی	۳	۳۸	شاغل	۸۴	روستایی
فرا	۸۰	۸۳	شاغل	۸۴	روستایی
وا	۹	۳	شاغل	۸۴	روستایی
نی	۳	۳۸	شاغل	۸۴	روستایی
جنسیت	مرد	زن	شغل	محل سکونت	وضعیت سکونت
فرا	۸۰	۸۳	شاغل	۸۴	روستایی
وا	۹	۳	شاغل	۸۴	روستایی
نی	۳	۳۸	شاغل	۸۴	روستایی
فرا	۸۰	۸۳	شاغل	۸۴	روستایی
وا	۹	۳	شاغل	۸۴	روستایی
نی	۳	۳۸	شاغل	۸۴	روستایی
جنسیت	مرد	زن	شغل	محل سکونت	وضعیت سکونت
فرا	۸۰	۸۳	شاغل	۸۴	روستایی
وا	۹	۳	شاغل	۸۴	روستایی
نی	۳	۳۸	شاغل	۸۴	روستایی
فرا	۸۰	۸۳	شاغل	۸۴	روستایی
وا	۹	۳	شاغل	۸۴	روستایی
نی	۳	۳۸	شاغل	۸۴	روستایی



افرادی که بالای ۵۵ سال سن دارند ۱۲ برابر افرادیست که کمتر از ۵۵ سال دارند.

با در نظر گرفتن متغیر تعداد کل موارد فوت شده به عنوان offset شاخص AIC و BIC برای هر دو مدل رگرسیونی پواسون و دو جمله‌ای منفی محاسبه شد، از آن جایی که درجه آزادی هر دو مدل برابر یک هست قادر به آزمون نسبت درستنمایی نیستیم.

اما با این وجود مشاهده می شود که مدل پواسن آکائیکه کمتری نسبت به مدل دو جمله‌ای منفی دارد بنابراین مدل پواسن به عنوان مدل بهینه انتخاب می شود هم چنین در نمودار Rootogram رسم شده برای این دو مدل تفاوت چندانی بین مدل رگرسیون پواسون و مدل رگرسیون دو جمله‌ای منفی مشاهده نشد، بنابراین انتخاب مدل ساده تر (پواسن) ارجحیت دارد.

در مرحله انتخاب متغیر مدل های تک متغیره با در نظر گرفتن توزیع پواسن و متغیر تعداد کل موارد فوت شده به عنوان offset بر روی داده ها برازش داده شد. سپس متغیرهای کمکی که مقدار p برای آن ها بیشتر از ۰/۲، به دست آمد، از مدل چندگانه حذف شدند.

با توجه به نتایج رگرسیون چند گانه پواسن ، مشخص شد که نسبت مرگ در اثر بیماری قلبی و عروقی در افرادی که بالای ۵۵ سال سن دارند تقریبا ۲ برابر افرادیست که کمتر از ۵۵ سال دارند. اما متغیرهای شغل همسر، وضعیت تحصیلات، نحوه سکونت تاثیر معناداری بر روی نسبت مرگ به دلیل بیماری قلبی و عروقی نداشتند.

با در نظر گرفتن متغیر زمان تا مرگ به عنوان offset مدل های دو جمله ای منفی و پواسن به داده ها بدون متغیر کمکی برازش داده شد. شاخص های AIC و BIC در جدول ۲ آورده شده است. هم چنین با استفاده از آماره آزمون نسبت درست نمایی مشخص شد مدل دو جمله ای منفی برازش بهتری بر روی داده ها دارد ($P < 0/001$). هم چنین در شکل ۱ نمودار Rootogram برای این دو مدل رسم شد که همانند آزمون نسبت درست نمایی نشان می دهد که مدل دو جمله ای منفی برازش بهتری روی داده ها دارد.

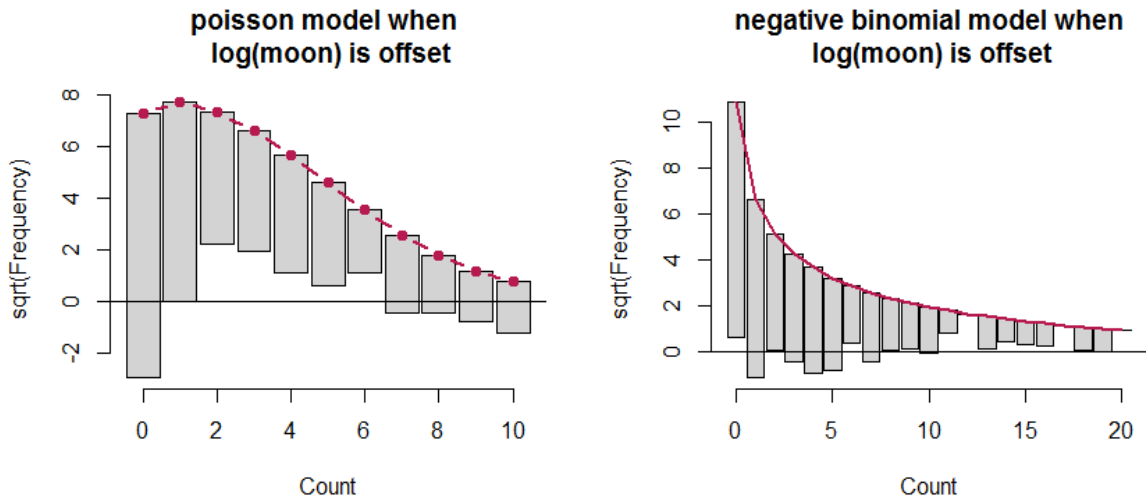
بنابراین در مرحله انتخاب متغیر مدل های تک متغیره با در نظر گرفتن توزیع دو جمله ای منفی و متغیر زمان تا مرگ در اثر بیماری قلبی و عروقی به عنوان offset بر روی داده ها برازش داده شد. سپس متغیرهای کمکی که مقدار p برای آن ها بیشتر از ۰/۲ بود از مدل چندگانه حذف شدند.

با توجه به نتایج رگرسیون چند گانه دو جمله ای منفی که در جدول ۳ مشخص شد نرخ مرگ در اثر بیماری قلبی و عروقی در افرادی که همسر آن ها خانه دار است نسبت به شاغل تفاوت معناداری ندارد. برای متغیر سطح تحصیلات می توان گفت نرخ مرگ در اثر بیماری قلبی و عروقی در افرادی که حداقل تحصیلات ابتدایی دارند ۶۵ درصد بیشتر از افراد بی سواد است. نرخ مرگ در اثر بیماری قلبی و عروقی برای افرادی که محل سکونت آن ها مرکز استان است، ۵۲ درصد بیشتر از سایر شهرستان ها است. نرخ مرگ در اثر بیماری قلبی و عروقی برای افرادی که شهری هستند تقریبا دو برابر روستایی ها است. هم چنین می توان گفت مرگ در اثر بیماری قلبی و عروقی در



جدول ۲: مقایسه شاخص های برازش مدل صفر وقتی offset متغیر زمان است

شاخص AIC	شاخص BIC	درجه آزادی	مدل های برازش شده
۱۸۸۴	۱۸۸۸	۱	پواسون
۱۲۳۸	۱۲۴۶	۲	دو جمله ای منفی



نمودار ۱: نمودار rootogram برای دو مدل پواسن و دو جمله ای منفی

جدول ۳: نتایج رگرسیون چندگانه مدل دو جمله ای منفی با حضور متغیرهای معنی دار

p	ZV	انحراف استاندارد	برآورد	
<۰/۰۰۱	-۴/۵۶۱	۰/۱۸۵۹	-۰/۸۴۸۷	عرض از مبدا
۰/۶۳۳۱	-۰/۴۷۷۴	۰/۱۶۰۵	-۰/۰۷۶۶۳	شغل (خانه دار نسبت به شاغل)
۰/۰۰۱۹۵۸	-۳/۰۹۷	۰/۱۶۳۱	-۰/۵۰۵	تحصیلات (بیسواد نسبت به باسواد)
۰/۰۰۹۷۴۴	۲/۵۸۵	۰/۱۶۲۴	۰/۴۱۹۹	محل سکونت (یزد نسبت به سایر شهرها)
<۰/۰۰۱	۳/۹۰۳	۰/۰۱۸۶۴	۰/۷۲۷۳	نحوه سکونت (شهری نسبت به روستایی)
<۰/۰۰۱	-۱۱/۲۳	۰/۲۲۱۸	-۲/۴۹	سن (زیر ۵۵ سال نسبت به بالای ۵۵ سال)

جدول ۴: مقایسه شاخص های برازش مدل صفر وقتی offset کل افراد فوت شده است

شاخص AIC	شاخص BIC	درجه آزادی	مدل های برازش شده
۷۴۵	۷۴۸/۷	۱	پواسون
۷۷۹/۷	۷۸۳/۴	۱	دو جمله ای منفی



بحث و نتیجه گیری

این مطالعه نگاه ویژه ای به متغیر *offset* در مدل بندی رگرسیون شمارشی دارد. این متغیر غالباً در مدلسازی رگرسیون شمارشی حذف می شود. از آن جایی که در رگرسیون شمارشی فرض بر آن است که تعداد پیشامدها در یک بازه زمانی یا مکانی مشخص را مورد بررسی قرار می دهیم، اگر پیشامد مورد بررسی برای افراد شرکت کننده در مطالعه، همگی در یک بازه زمانی یا مکانی باشد نیازی به ورود این متغیر نیست. اما اغلب در عمل، فراوانی پیشامد ثبت شده در بازه های زمانی یا مکانی متفاوتی روی می دهند. بنابراین توجه به *offset* بسیار ضروری است. علاوه بر این نوع متغیر *offset* می تواند زمینه مدل سازی را تحت تأثیر قرار دهد. اگر زمان تا پیشامد مورد نظر به عنوان *offset* در نظر گرفته شود در این صورت پاسخ به جای تعداد، نرخ پیشامد است و اگر تعداد کل افراد را به عنوان *offset* در نظر بگیریم پاسخ درصد پیشامد است. این موضوع باید در تفسیر ضرایب در نظر گرفته شود.

بنابراین مدل های دو جمله ای منفی و پواسنی با در نظر گرفتن زمان و تعداد کل افراد به عنوان *offset* بر روی داده ها پردازش شد. هنگامی که متغیر زمان تا پیشامد به عنوان *offset* در نظر گرفته شد مدل دو جمله ای منفی از نظر *AIC* برآزش بهتری داشت. این موضوع به دلیل وجود بیش پراکنش در توزیع نرخ مرگ و میر در اثر بیماری های قلبی و عروقی است. اما هنگامی که تعداد کل افراد به عنوان *offset* در نظر گرفته شد مدل پواسنی بهتر از دو جمله ای منفی برآزش شد.

سن در همه تحلیل ها، از نظر آماری معنی دار شده است. در مدل رگرسیونی دو جمله ای منفی نرخ مرگ بر اثر بیماری های قلبی

و عروقی در افرادی که بالای ۵۵ سال سن دارند ۱۲ برابر افرادیست که کمتر از ۵۵ سال دارند. و در مدل رگرسیونی پواسن که نسبت مرگ بر اثر بیماری های قلبی و عروقی در افرادی که بالای ۵۵ سال سن دارند تقریباً ۲ برابر افرادیست که کمتر از ۵۵ سال دارند. که مطابق با مطالعه ای است که عامری و همکاران در سال ۲۰۱۴ روی عوامل تاثیر گذار بر بیماری های قلبی انجام دادند سن را به دو قسمت زیر ۴۵ سال و بالای آن در نظر گرفتند. و سن بالا از عوامل مهم و تاثیر گذار بر بیماری های دیگر در کنار سایر عوامل شناخته شد (۲۱). اما برخلاف نتایج ما در مطالعه Judith و همکاران در سال ۲۰۱۳ مرگ و میر ناشی از بیماری های قلبی را بر اساس کشور، مذهب و سن بررسی کردند و دریافتند که سن به تنهایی از عوامل موثر بر مرگ بر اثر بیماری های قلبی نیست اما در کنار عواملی هم چون وضعیت اقتصادی میتواند تاثیر گذار باشد (۲۲).

متغیر جنسیت در هیچ یک از تحلیل ها، از نظر آماری معنی دار نشده است و در هیچ یک از مدل های رگرسیون شمارشی آن را وارد مدل بندی نکردیم. در بسیاری از مطالعاتی که در این زمینه انجام شد عامل جنس معنی دار نشده است مانند مطالعه Furer و همکاران در سال ۲۰۱۸ در مطالعه ای شاخص *BMI* را با مرگ میر ناشی از بیماری های قلبی را بررسی کردند و دریافتند که به جز زنانی که دچار کمبود وزن شدید هستند مرگ و میر بیماری های قلبی در زنان و مردان تفاوت معناداری ندارد (۲۳). اما برخلاف نتایج ما در مطالعه Sun و همکاران در سال ۲۰۱۸ در یک متاآنالیز که در رابطه دیابت و بیماری های قلبی و سایر مرگ و میرها روی ۶۸ مطالعه بود دریافتند که خطر مرگ و میر در اثر بیماری های قلبی در زنان بیشتر از مردان است (۲۴).



شهر است (۲۷).

نتایج نشان داد در صورتی زمان وقوع مرگ به عنوان متغیر offset در نظر گرفته شود، مدل رگرسیون دو جمله منفی برای نشان دادن عوامل موثر بر مرگ به علت بیماری قلبی و عروقی با توجه به معیار AIC و BIC کارآتر می باشد. در این حالت مشخص شد نرخ مرگ در اثر بیماری قلبی و عروقی برای جنس و شغل تفاوت معناداری ندارد. نرخ مرگ در اثر بیماری قلبی و عروقی با استفاده از رگرسیون دو جمله ای منفی برای متغیرهای تحصیلات، محل سکونت، نوع سکونت و سن معنی دار شد هم چنین اگر تعداد کل موارد فوت شده به عنوان offset در نظر گرفته شود، مدل رگرسیون پواسون با توجه به معیارهای AIC, BIC کارآتر از مدل دو جمله ای منفی می باشد که در این حالت نسبت مرگ به علت بیماری قلبی و عروقی برای متغیر سن معنی دار و برای متغیرهای شغل، تحصیلات، نحوه سکونت و محل سکونت معنی دار نشد.

تضاد منافع

نویسندگان این مقاله اعلام می دارند که در این مقاله هیچ گونه تضاد منافی وجود ندارد.

تشکر و قدر دانی

از کلیه کسانی که در انجام این تحقیق ما را یاری رساندند کمال تشکر و قدر دانی را داریم.

References

1-Abubakar I, Tillmann T, Banerjee A. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2015;385(9963):117-71.

متغیر محل سکونت در مدل دو جمله ای منفی معنادار شد به این صورت که نرخ مرگ و میر در اثر بیماری های قلبی و عروقی برای افرادی که محل سکونت آن ها مرکز استان است، ۵۲ درصد بیشتر از سایر شهرستان ها است. اما این متغیر در مدل رگرسیون پواسن از نظر آماری معنی دار نبود. مطابق است با مطالعه عبدلهی و همکاران در سال ۲۰۱۲ که در یک مطالعه توصیفی-تحلیلی روی ۵۴۴۴ نفر دچار بیماری قلبی بودند انجام دادند و دریافتند بین این بیماری و محل سکونت رابطه معناداری وجود ندارد (۲۵).

و در آخر متغیر وضعیت سکونت در مدل دو جمله ای منفی معنادار شد به این صورت که نرخ مرگ به علت بیماری های قلبی و عروقی برای افرادی که شهری هستند تقریباً دو برابر روستایی ها است. اما این متغیر در مدل رگرسیون پواسن از نظر آماری معنی دار نبود. در مطالعه مقطعی درویش پور و همکاران در سال ۲۰۱۶ که روی ۱۰۶ نفر از بیماران قلبی مراجعه کننده در رشت انجام دادند، محل سکونت به عنوان شاخص معنادار روی بیماری های قلبی عروقی شناخته شد و کسانی در شهر زندگی می کردند در مواجهه با بیماری های قلبی بودند (۲۶). اما در مطالعه آصف زاده و همکاران در سال ۲۰۱۳ که در مورد فوت شدگان بر اثر بیماری های قلبی و عروقی در شهر قزوین انجام دادند دریافتند که مرگ و میر در روستا دوبرابر مرگ و میر در



- 2-Kazemnejad A, Faramand S, Zayeri F, Salehi M, Yazdani Cherati J. Preparing the geographical maps of the relative death rate out of vasco-cardiac diseases in cities of the mazandaran province in 2008. *Journal of Mazandaran University of Medical Sciences*.2012;22(94):63-9.
- 3-Forouzanfar MH, Sepanlou SG, Shahrzaz S, BESc PN, Pourmalek F, Lozano R, et al. Evaluating causes of death and morbidity in Iran, global burden of diseases, injuries, and risk factors study 2010. *Archives of Iranian medicine*. 2010; 304:(5)17;14.
- 4-Kontis V, Mathers CD, Bonita R, Stevens GA, Rehm J, Shield KD, et al. Regional contributions of six preventable risk factors to achieving the 25× 25 non-communicable disease mortality reduction target: a modelling study. *The Lancet Global Health*. 2015;3(12):e746-e57.
- 5-Rashidi M, Ghias M, Ramesht MH. Geographical Epidemiology of Death Due to Cardiovascular Diseases in Isfahan Povince, Iran. *Journal of Isfahan Medical School*. 2011;29(125).
- 6-Rezaeian M, Dehdarinejad A, Esmaili Nadimi A ,Tabatabaie S. Geographical epidemiology of deaths due to cardiovascular diseases in counties of kerman province. *Iranian Journal of Epidemiology*. 2008;4(1):35-41.
- 7-Woodward M. Cardiovascular disease and the female disadvantage. *International journal of environmental research and public health*. 2019;16(7):1165.
- 8-Mertler CA, Reinhart RV. *Advanced and multivariate statistical methods: Practical application and interpretation*: Routledge; 2016.
- 9-Kulkarni VG. *Modeling and analysis of stochastic systems*: Chapman and Hall/CRC; 2016.
- 10-Imprialou M-IM, Quddus M, Pitfield DE, Lord D. Re-visiting crash–speed relationships: A new perspective in crash modelling. *Accident Analysis & Prevention*. 2016;86:173-85.
- 11-Coggins BL. *Comparing models for fitting count data*: University of South Carolina; 2009.
- 12-Chow N, Steenhard D. A flexible count data regression model using SAS Proc nlmixed. *SAS Global Forum: Statistics and Data Analysis*; 2009: Citeseer.
- 13-Cox DR. *Analysis of binary data*: Routledge; 2018.
- 14-Ardiles LG, Tadano YS, Costa S, Urbina V, Capucim MN, da Silva I, et al. Negative Binomial regression model for analysis of the relationship between hospitalization and air pollution. *Atmospheric Pollution Research*.2018;9(2):333-41.



- 15-Waenerlund AK, Mosquera PA, Gustafsson PE, San Sebastian M. Trends in educational and income inequalities in cardiovascular morbidity in middle age in Northern Sweden 1993–2010. *Scandinavian journal of public health*. 2018;1403494818790406.
- 16-Correia JC, Braillard O, Combescure C, Gerstel E, Spechbach H. Admission rates in emergency departments in Geneva during tennis broadcasting: a retrospective study. *BMC emergency medicine*. 2018;18(1):56.
- 17-Joe H, Zhu R. Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*. 2005;47(2):219-29.
- 18-Chang LY. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. *Safety science*. 2005;43(8):541-57.
- 19-Sharifirad G, Mohebbi S, Matlabi M. The relationship of physical activity in middle age and cardiovascular problems in old age in retired people in Isfahan, 2006. *The Horizon of Medical Sciences*. 2007;13(2):57-63.
- 20-Dogra S, Ardern CI, Baker J. The relationship between age of asthma onset and cardiovascular disease in Canadians. *Journal of Asthma*. 2007;44(10):849-54.
- 21-Ameri H, Alizadeh S, Barzegari A. Identification of influencing factors for heart attack in diabetic patients using C & R algorithm. 2014.
- 22-Finegold JA, Asaria P, Francis DP. Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations. *International journal of cardiology*. 2013;168(2):934-45.
- 23-Furer A, Afek A, Orr O, Gershovitz L, Rabbi ML, Derazne E, et al. Sex-specific associations between adolescent categories of BMI with cardiovascular and non-cardiovascular mortality in midlife. *Cardiovascular diabetology*. 2018;17(1):80.
- 24-Sun LY, Tu JV, Sherrard H, Rodger N, Coutinho T, Turek M, et al. Sex-Specific trends in incidence and mortality for urban and rural ambulatory patients with heart failure in eastern Ontario from 1994 to 2013. *Journal of cardiac failure*. 2018;24(9):568-74.
- 25-Abdollahi AA, Hoseini SA, Salehi A, Behnampour N, Abasi A. Coronary artery lesions and some of its related factors in more than 5000 patients in kosar Angiography Center (Golestan



Province) from 2007 to 2009. Scientific Journal of Kurdistan University of Medical Sciences. 2012;17(1):18-24.

26-Darvishpour A, Javadi Pashaki N, Salari A, Taleshan Nejad M, Barari F. Comparing the quality of life in patients with cardiovascular diseases before and after coronary angioplasty. Journal of Mazandaran University of Medical Sciences. 2016;26(137):206-10.

27-Asefzadeh S, Alikhani S, Javadi H. Socio-economic status and mortality from cardiovascular diseases in Qazvin. 2013;16(4):40-46.