



دانشگاه علوم پزشکی کرمان

دانشکده بهداشت

پایان نامه مقطع کارشناسی ارشد رشته آمار زیستی

عنوان:

مقایسه رگرسیون دوجمله‌ای منفی صفرمتورم بیزی باغیربیزی برداده‌های

شمارشی شناخت افرادالکلی

توسط: سپیده روحی

استاد راهنما: دکتر عباس بهرامپور

استاد مشاور: دکتر محمدرضا بانشی

سال تحصیلی ۱۳۹۵-۱۳۹۶



Kerman University of Medical Sciences

Faculty of Medicine

In Partial Fulfillment of the Requirements for the Degree MSc

Titel:

**Comparing Bayesian regression and classic zero-inflated negative binomial
on size estimation of people who use alcohol**

By:

Sepideh Roohi

Supervisor:

Dr bas Bahrampour

Advisor:

Dr Mohammad Reza Baneshi

Year:

August 2017

مقدمه دو اهداف: معتادین و مصرف کنندگان الکل (بخصوص در جوامعی که مصرف الکل و موادمخدر در آن غیر-

مجاز می باشد) گروه های پنهان جامعه هستند که محققین برای بدست آوردن اطلاعاتشان با مشکلات عدیده ای روبرو

می شوند. از آنجایی که این افراد در اجتماع از مقبولیت پایینی برخوردار هستند و سعی در مخفی نمودن مشکلات خود را

دارند، پاسخ به این سوال که شخص چند نفر از افراد متعلق به این گروه های پنهان را می شناسد همواره با صفرهای زیادی

روبرو می شود. در دنیای امروز بالاخص در سالهای اخیر اکثر علوم با داده های شمارشی همراه با صفر زیاد مواجه هستند.

این قبیل داده ها بیشتر در رشته های پزشکی، بهداشت محیط، کشاورزی و غیره مشاهده می شوند. بطور کلی مدلی که برای

آنها در نظر گرفته شده است مدل رگرسیونی پواسن متورم در صفر است که در عمل با مشکلاتی مانند بیش پراکنش مواجه

می گردد که مدل دو جمله ای منفی متورم در صفر مناسب تر است. بنابراین سوالی که باید در نظر گرفت این است که کدام از

دو روش بیزی و غیربیزی (کلاسیک) در این پژوهش مناسب می باشد و سوال دوم این است که آیا متغیرهایی مانند سن،

تحصیلات و وضعیت تاهل بر شناخت افراد از گروه های پنهان تاثیرگذار است.

روشها: در این مطالعه که بر روی ۷۰۰ نفر در دو استان کرمان ۳۲۸ نفر (۴۶/۹ %) و استان فارس با ۳۷۲ نفر (

۵۳/۱ %) صورت گرفته است و از آنها پرسیده شده چه تعداد از افراد متعلق به گروه های پنهان را می شناسد. متغیرهای

وابسته در مطالعه شناخت تعداد افراد الکل، شناخت تعداد افراد مصرف کننده متادون و متغیرهای مستقل استان، جنسیت،

تحصیلات و وضعیت تاهل است. جهت برازش مدل رگرسیون دو جمله ای منفی متورم در صفر از دو روش بیزی و غیربیزی

استفاده شده است و پس از آن این دو روش با هم مقایسه شده اند. نتایج روش بیزی را در نرم افزار OpenBUGS و از

طریق کدهای مربوطه در R و نتایج روش کلاسیک نیز در نرم افزار R استخراج شد. و براساس فواصل اطمینان احتمالی و خطای استاندارد مقایسه شدند.

یافته ها: در شناخت افراد مصرف کننده الککل و متادون، درصد صفر به ترتیب ۳۳/۹٪ و ۴۹/۱٪ بوده است، و با توجه به وجود بیش پراکنش در این مطالعه (عدم برابری میانگین و پراش) مدل بکار رفته دوجمله ای منفی متورم در صفر بوده است. در روش کلاسیک چهار متغیر استان، جنسیت، گروه های سنی و تحصیلات بر روی شناخت افراد الکلی تاثیرگذار بوده اند و در متغیر پاسخ متادون دو متغیر استان و تحصیلات معنادار شده اند. اما در روش بیزی تنها سه متغیر جنسیت، سن و تحصیلات برای متغیر پاسخ الککل و دو متغیر استان و تحصیلات برای متغیر پاسخ متادون در هر دو روش معنادار شدند.

بحث و نتیجه گیری. در نهایت پس از تجزیه و تحلیل مشخص شد که روش بیزی برای داده های شناخت افراد الکلی عملکرد بهتری نسبت به روش کلاسیک دارد. متغیرهای تاثیرگذار در روش کلاسیک استان، تحصیلات، سن و جنس و در روش بیزی متغیرهای تحصیلات، سن و جنس است، اما در داده های شناخت افراد مصرف کننده متادون عملکرد روش بیزی با روش کلاسیک یکسان بوده است.

واژگان کلیدی: بیزی، زنجیره مارکوف مونت کارلو، رگرسیون دوجمله ای منفی متورم در صفر، داده های شمارشی

Abstract

Background and Objective: Nowadays we have some data in different sciences which number of zeros is more than expected, such data are called zero-inflated. which can be modeled by regressions for count data. Many researches have been conducted in the field of classical method on count data. Most of Bayesian analysis which are conducted for these data used zero-inflated Poisson regression. Therefore the main purpose of this research is comparison of Bayesian and classic approaches in regression of zero-inflated negative binomial on data for determining the size estimation of people who have used alcohol more than once in last year.

Method: This research had been in two provinces of Fars and Kerman in 2011, sample size of each province was formed proportional to people of that province and totally the calculated sample size was 700. Zero-inflated negative binomial regression was fitted to The data in two Bayesian and classical methods and then two methods have been compared. Results of Bayesian method were extracted in OpenBUGS software and through related codes in R and results of classical method were extracted in R software too.

findings: After fitting classical method, variables of province, gender, age groups and education had been effective on identifying number of alcoholics, but in Bayesian method three variables of gender, age groups and education have become significant. In this research it was specified that obtained probability intervals from Bayesian method are much wider than classical method.

conclusion: Results of this research indicate that Bayesian method has better function than the classic.

Keywords: Bayesian, MCMC, Zero-inflated negative binomial, Count data.