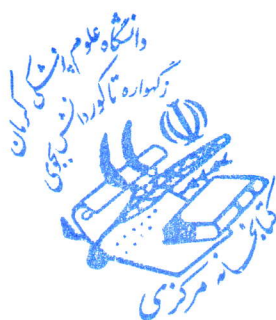




کاربرد مدل‌های رگرسیون بوم‌شناختی در مدلسازی داده‌های گروهی و تعیین

عوامل موثر در مصرف الک

پایان نامه مقطع کارشناسی ارشد



توسط:

فاطمه مهدی‌زاده

استاد راهنما:

دکتر محمدرضا بانیشی

استاد مشاور:

دکتر علی‌اکبر حق‌دوست

دانشگاه علوم پزشکی کرمان

سال تحصیلی ۹۵-۹۴



**Application of Ecological Regression Model in Analysis of Aggregate Data and the
Factors Related to Alcohol**

A Thesis

Presented to

The Graduate Studies

By:

Fatemeh Mahdizadeh

Supervisor:

Dr.Mohammad Reza Baneshi

Advisor:

Dr.AliAkbar Haghdoost

Kerman University of Medical sciences

June 2016



مقدمه و هدف:

در اکثر پژوهش‌های آماری داده‌ها به صورت فردی می‌باشند. بدین معنی که اطلاعات متغیر وابسته و متغیرهای مستقل برای تک تک افراد موجود است. در برخی موارد اطلاعات فردی در دسترس نیست در این مطالعات از داده‌های مشاهده‌ای جمع‌آوری شده روی گروه‌هایی که بوسیله مناطق جغرافیایی دسته بندی شده اند استفاده می‌شود. در این نوع داده‌ها معمولاً اطلاعات منابع مختلف در سطح گروه با هم ترکیب شده و به صورت خلاصه شده بیان می‌شود. این نوع داده‌ها کاربردهای زیادی در سیستم سلامت، آزمایشگاهها و غیره دارند. برای تجزیه و تحلیل داده‌های گروهی نیاز به استفاده از مدل رگرسیون بوم‌شناختی است که از داده‌ها در سطح گروهی و نه فردی برای برازش مدل رگرسیونی استفاده می‌کند. به علت جدید بودن این موضوع پژوهش‌های زیادی در این زمینه انجام نشده است اما در اندک مطالعات انجام شده اکثراً توصیه به استفاده از رگرسیون بوم‌شناختی شده است. هدف ما بررسی کاربرد مطالعه بوم‌شناختی در داده‌های شناخت مصرف کنندگان الکلی و مقایسه آن با روشهای آتالیز معمول است تا ببینیم چه میزان بهبودی در برآورد حاصل می‌شود.

روش کار:

در این پژوهش از داده‌های بانک اطلاعات کشوری که ۱۲۰۰۰ داده در مجموع ۳۱ استان کشور جمع‌آوری شده است استفاده شده است. برای جمع‌آوری اطلاعات آن از دو پرسشنامه استفاده شده است. پرسشگری در خیابان و از عالیرین پیاده بالای ۱۸ سال صورت گرفته و سعی شده از تمام رده‌های سنی و طبقات اجتماعی وارد مطالعه شوند. علاوه

بر این داده‌ها چهار متغیر گروهی نیز از مرکز آمار ایران جمع‌آوری شده است. در این پژوهش با دو نوع داده روبرو هستیم داده‌های فردی و گروهی. متغیر پاسخ شناخت مصرف‌کنندگان الکل است که دو حالتی است یا فرد مصرف‌کننده الکل را می‌شناسد (کد ۱) و یا نمی‌شناسد (کد ۰). متغیرهای سن، جنس، شناخت مصرف‌کنندگان شیره تریاک، شناخت مصرف‌کنندگان اکستازی، شناخت مصرف‌کنندگان حشیش و شناخت مصرف‌کنندگان سایر مواد متغیرهای مستقل ما هستند که سن کمی و سایر متغیرها کیفی دو حالتی می‌باشد یعنی مردان کد ۱ و زنان کد ۰ و در مورد سایر متغیرها همانند متغیر پاسخ شناخت مصرف‌کنندگان الکل کد ۱ را به شناخت مصرف‌کننده و کد ۰ را به عدم شناخت مصرف‌کننده نسبت داده‌ایم لذا به داده‌ها در حالت فردی به علت دو حالتی بودن پاسخ رگرسیون لجستیک را برازش دادیم سپس داده‌ها را با توجه به کمی یا کیفی بودن آن به ترتیب با میانگین و نسبت در آن منطقه خلاصه کردیم و سپس رگرسیون بوم‌شناختی را برازش دادیم. در ادامه متغیرهای که از مرکز آمار ایران استخراج کردیم را به داده‌ها اضافه کرده و رگرسیون بوم‌شناختی را برازش دادیم. آنالیزها با نرم‌افزارهای R و SPSS انجام شده‌اند.

یافته‌ها:

مدل رگرسیون لجستیک به عنوان مدل مطلوب در نظر گرفته شده و با نتایج رگرسیون بوم‌شناختی مقایسه شده است. با وجود کاهش حجم نمونه از ۱۲۰۰۰ به ۳۱ نتایج مدل لجستیک با مدل بوم‌شناختی تفاوت زیادی ندارند. در مدل لجستیک همه متغیرها معنی دار شدند از طرفی در مدل بوم‌شناختی همه متغیرها به جز دو متغیر سن و شناخت مصرف‌کنندگان حشیش معنی دار شدند. مثلاً در مدل لجستیک OR برای متغیر جنس برابر با ۱,۳۵ می‌باشد یعنی مردان ۳۵ درصد بیشتر از زنان مصرف‌کننده الکل می‌شناسند؛ در مورد متغیر سن OR در مدل لجستیک برابر با ۰,۹۸ می‌باشد یعنی شانس شناخت مصرف‌کنندگان الکل با افزایش سن ۲ درصد کاهش می‌یابد؛ در مورد سایر متغیرها یعنی شناخت مصرف‌کنندگان شیره تریاک، شناخت مصرف‌کنندگان اکستازی، شناخت مصرف‌کنندگان حشیش و

شناخت مصرف کنندگان سایر مواد میزان OR در مدل لجستیک به ترتیب برابر با ۱,۸۵، ۱,۹، ۲,۷۶ و ۱,۷۴ می باشد یعنی شانس شناخت مصرف کنندگان الکل در کسانی که مصرف کنندگان شیره تریاک، اکستازی، حشیش و سایر مواد را می شناسند به ترتیب ۸۵، ۹۰، ۲,۷۶ و ۷۴ درصد بیشتر از کسانی است که این مصرف کنندگان را نمی شناسند.

نتیجه گیری:

از این مطالعه نتیجه می شود برای انجام مطالعاتی که در سطح منطقه ای یا جغرافیایی هستند به جای استفاده از روش های کلاسیک معمول برای جمع آوری و آنالیز داده ها می توان از استنباط بوم شناختی استفاده کرد زیرا داده ها در سطح منطقه ای بوده و از طرفی برای صرفه جویی در زمان به جای جمع آوری داده ها در سطح فردی، داده ها را به صورت انباشته یا گروهی جمع آوری می شوند لذا هم از نظر مالی و هم از نظر زمانی پیشنهاد استفاده از این نوع آنالیز را می دهیم.

کلمات کلیدی:

شناخت مصرف کنندگان الکل، داده های گروهی، رگرسیون بوم شناختی، داده های فردی، رگرسیون لجستیک

Abstract

In majority of data sets, data are gathered at individual level. This means that information of outcome and independent variables are available for all subjects. Sometimes data are not available at individual level, but at province or national level. In social sciences, usually observed data are routinely collected and grouped by geographical areas. Such data sets are known as ecological data. For analysis of group data needs to ecological regression model. Our goal is to study the application of ecological study in data consumption of alcohol and comparison with logistic regression to see how much improvement can be achieved in the estimation

Methods

We used a national data set that was collected to estimate prevalence of several risky behaviors among Iranian population. In the original study, a sample of about 500 participants was recruited at each province. After data cleaning the sample size was about 12000. Respondents were asked whether they know anybody who had drunk alcohol at least once in the past year. This variable was considered as the dependent variable. In addition to that, the following aggregated variables were extracted for each province from official statistics data: urbanization rate, literacy percentage, household size and share of the country's population. Then all variables were combined at province level. The percentage of the positive answers to the response variable was considered as the dependent variable. For continuous and binary variables, mean and proportion at each province was calculated respectively. Ecological regression model was fitted. Here, the sample size was 31. The two models fitted once before and once after added aggregate variables. Statistical analyzes were performed with software SPSS and R.

Results

The logistic regression model is considered as a preferred model and compared with the results of ecological regression. Despite the decrease in sample size from 12,000 to 31 results ecological regression closed to logistic regression. So that in model with aggregate variables just 2 and in model without aggregate variables just 1 variable are not matching in terms of direction of OR.

On the other hand, all variables were significant in logistic model and in ecological Model all variables except age and know gras users were significant. For example, the logistic model for sex variable OR is equal to 1.35 that is 35 percent more men than women are known alcohol user; about age variable OR equal to 0.98 that is odds of know alcohol user 2 percent decrease while age increase. In other variables that is know opium, ecstasy, gras and other material OR in the logistic model are 1.85, 1.9, 2.76 and 1.74, that is know alcohol user in opium, ecstasy, gras and other material are 85, 90, 2.76, 74 percent more than those who do not know alcohol user and in ecological models 1.89, 4.3, 2.13 and 2.63, respectively.

Conclusion

As a result of this study on the area or geography level instead of usual classical methods for collecting and analyzing data can be used to ecological inference because the data are in area level and can save time. so, we recommend the use of this type of analysis

Keywords: know alcohol user, aggregate data, ecological regression, individual data, logistic regression