



دانشگاه علوم پزشکی کرمان

دانشکده بهداشت

پایان نامه مقطع کارشناسی ارشد آمار زیستی

عنوان :

مقایسه مدل های رگرسیون لجستیک فازی و رگرسیون لجستیک برای پیش بینی بقای بیماران مبتلا  
به سرطان پستان

توسط : رویا نیک بخت

استاد راهنما : دکتر عباس بهرامپور

سال تحصیلی : ۱۳۹۲-۱۳۹۳

Comparison between fuzzy logistic regression and logistic regression models to  
predict survival of patients with breast cancer

A Thesis  
Presented to  
The Graduate Studies

By  
**Roya Nikbakht**

In Partial Fulfillment  
Of the requirements for the Degree  
Master of Science in:  
**Biostatistics**

**Kerman University of Medical Sciences**  
**Jun 2014**



## چکیده :

**مقدمه و هدف :** در سال‌های اخیر، توجه قابل ملاحظه‌ای به مدل‌های آماری برای طبقه‌بندی داده‌های پزشکی با توجه به بیماری‌های مختلف و پیامدهای آن‌ها شده است. رگرسیون لجستیک فازی در مواقعی که داده‌ها با ابهام همراه باشند مورد استفاده قرار می‌گیرد. هدف از این مطالعه، مقایسه دو مدل آماری رگرسیون لجستیک فازی و رگرسیون لجستیک برای پیش‌بینی مدت زمان بقای بیماران مبتلا به سرطان پستان است.

**مواد و روش‌ها :** داده‌های موجود توسط مرکز ثبت سرطان کرمان طی سال‌های ۸۶-۱۳۷۹ جمع‌آوری گردیده است. جهت پیش‌بینی بقای هر بیمار، دو مدل رگرسیون لجستیک و رگرسیون لجستیک فازی استفاده شد. داده‌های مورد استفاده در این مطالعه شامل ۹۳۴ بیمار مبتلا به سرطان پستان می‌باشد. از ۵ نمونه‌ی ۷۱ تایی و یک نمونه‌ی ۱۶ تایی جهت انجام مدلسازی استفاده گردید. رده‌ی سنی بیماران سرطانی در این مطالعه ۱۵ تا ۸۵ سال بود. برای تعیین نیکویی برازش دو مدل همچنین مقایسه‌ی دو مدل از شاخص‌های حساسیت، ویژگی و سطح زیر منحنی مشخصه عملکرد (ROC) استفاده گردید.

**یافته‌ها :** در این مطالعه، حساسیت دو مدل رگرسیون لجستیک فازی و رگرسیون لجستیک به ترتیب  $(0.63 و 0.68)$ ،  $(0.62 و 0.68)$ ،  $(0.37 و 0.68)$ ،  $(0.73 و 0.73)$ ،  $(0.25 و 0.67)$ ،  $(0.66 و 1)$  و ویژگی در دو مدل رگرسیون لجستیک فازی و رگرسیون لجستیک به ترتیب  $(0.63 و 0.81)$ ،  $(0.63 و 0.73)$ ،  $(0.63 و 0.63)$ ،  $(0.27 و 0.81)$ ،  $(0.82 و 0.64)$  و  $(0.90 و 0.81)$  به دست آمد. همچنین سطح زیر منحنی راک برای مدل لجستیک فازی و رگرسیون لجستیک به ترتیب  $(0.58 و 0.81)$ ،  $(0.60 و 0.73)$ ،  $(0.53 و 0.76)$ ،  $(0.54 و 0.82)$ ،  $(0.63 و 0.78)$  و  $(0.81 و 0.91)$  بود.

**نتیجه‌گیری :** یافته‌ها نشان می‌دهند که تقریباً در اکثر نمونه‌ها حساسیت و ویژگی در مدل رگرسیون لجستیک بیشتر است.

**واژگان کلیدی :** رگرسیون لجستیک، رگرسیون لجستیک فازی، پیش‌بینی بقا، سرطان پستان

## **Abstract:**

**Background and objective:** In the recent years, considerable attention has been to the statistical models for classification of medical data according to various diseases and its outcomes. Fuzzy logistic regression measures the association between vague variables. In current study, we applied this model for prediction of survival time in patients with breast cancer in comparison to logistic regression model.

**Method:** We used the registered data by Cancer Registry Center in Kerman during 2000 - 2007. The fuzzy logistic regression and logistic regression were fitted on data for predicting 5-year survival of patients with breast cancer. We took a sample with sample sizes of 16 and 5 samples with sample size of 71 from 924 patients by Minitab software. The models were compared regarding sensitivity, specificity and The area under ROC curve (AUC).

## **Result:**

In this study, the sensitivity of fuzzy logistic regression and logistic regression models were (0.63, 0.68), (0.62, 0.68), (0.37, 0.68), (0.73, 0.73), (0.25, 0.67) and (0.66, 1) respectively and specificity of two models were (0.63, 0.81), (0.63, 0.73), (0.63, 0.63), (0.27, 0.81), (0.64, 0.82) and (0.81, 0.90). Also the area under curve ROC for two models was (0.58, 0.81), (0.60, 0.73), (0.53, 0.76), (0.54, 0.82), (0.63, 0.78) and (0.81, 0.91).

## **Conclusion:**

Our results showed that the sensitivity and specificity in logistic regression model is greater than fuzzy logistic regression in most samples.