View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Simorgh Research Repository

دانشگاه علوم پزشکی

و خدمات‌بهداشتی‌درمانی‌کرمان

دانشکده بهداشت

پایان نامه مقطع دکتری تخصصی (Ph.D.) آمار زیستی

عنوان :

به کارگیری مدل جنگل بقای تصادفی در داده های با بعد کم (low dimensional) و رویداد کم (few event) و مقایسه آن با تحلیل رگرسیون کاکس تحت روش های رگرسیونی جریمه شده و مدل شفایافته پارامتریک بیزی

توسط :

شیده رفعتی

اساتید راهنما :

دکتر عباس بهرامپور | دکتر محمدرضا بانشی

شماره پایان نامه: سال تحصیلی : ۱۳۹۹-۱۳۹۸

فهرست مطالب **صفحه**

فهرست جداول **صفحه**

فهرست اشکال

صفحه

فهرست پیوست‌ها

عنوان **صفحه**

چکیده

مقدمه و اهداف: امروزه با پیشرفت های چشم گیر در تمام حوزه های علم پزشکی، نسبت قابل توجهی از بیماران مبتلا به انواع مختلف بیماری بقای طولانی مدت دارند. پس در بسیاری از موارد با داده های بقایی مواجه هستیم که تعداد بیمارانی که مرگ بر اثر بیماری مورد نظر را تجربه می کنند، نسبت به موارد سانسور شده خیلی کمتر است. از طرف دیگر، با توجه به اینکه در بسیاری از تحقیقات با تعداد زیادی از متغیرهای مستقل روبرو هستیم لذا باید نسبت تعداد رویداد ها به تعداد متغیرهای مستقل یا به عبارت بهتر مقدار Event per Variable(EPV) مورد توجه قرار گیرد. براساس مطالعات شبیه سازی انجام شده، EPV بین ۱۰ تا ۲۰ توصیه شده است. وقتی EPV کم است، استفاده از مدل های کلاسیک همچون کاکس پیشنهاد نمی شود زیرا ضرایبی که محاسبه می شوند قابل اعتماد نمی باشند.

در داده های با بعد بالا، داده ای که تعداد متغیرهای پیش بین آن بسیار بیشتر از تعداد مشاهدات است، همواره EPV کم است. به همین جهت در صورت مواجهه با چنین داده ای بدون چک کردن مقدار EPV مستقیما با راه کارهایی همچون استفاده از روش های مجازات شده، از واریانس زیاد مدل های کلاسیک کاسته می شود. اما در مورد داده با بعد کم، یعنی داده ای که تعداد متغیرهای پیش بین کمتر از تعداد مشاهدات باشد، یا مقدار EPV مطلوب است و استفاده از مدل های کلاسیک بلامانع است یا مقدار EPV کم است و در نتیجه استفاده از مدل کلاسیک برای این داده مسئله ساز خواهد بود بنابراین، برای غلبه بر مدل بیش برازش یافته کلاسیک باید راهکاری اتخاذ شود تا واریانس مدل بهبود یابد.

پس نظر به اهمیت شاخص EPV و از آن جا که در مطالعات معدودی داده های با بعد کم و EPV کم مورد ارزیابی قرار گرفتند، در این پژوهش به آن پرداخته شده است و با توجه به ماهیت داده مورد استفاده، به طور خاص، این پژوهش با هدف مقایسه عملکرد مدل جنگل بقای تصادفی، مدل کاکس تحت روش های مجازات شده و مدل شفایافته بیزی به عنوان سه رویکرد متفاوت ناپارامتری، نیمه پارامتری و پارامتری در داده های با بعد کم و EPV کم به انجام رسیده است.

روش پژوهش: در این پژوهش ۲۵۲ بیمار دیالیزی مورد بررسی قرار گرفتند. به دلیل برخورداری از۳۵ رویداد مرگ و ۱۹متغیر، مقدار EPV تقریبا ۲ محاسبه گردید. این داده به طور تصادفی به دو مجموعه آزمودنی و آموزشی تقسیم شد و این عمل ۵۰۰ بار تکرار گردید. مدل جنگل بقای تصادفی و مدل های کاکس تحت روش های مجازات شده (لا سو، ریج و الا ستیک نت)به عنوان اولین و دومین راهکار پیشنهاد شده در داده با بعد کم و EPV کم، برای داده های آموزشی اجرا شدند و ارزیابی دقت پیش بینی این مدل ها در داده های آزمودنی با استفاده از شاخص های calibration slope(CS)، C-index و Brier Score (BS) انجام شد. بعلاوه، از آنجا که در داده مورد پژوهش تمامی رویداد ها در ابتدای زمان پیگیری رخ داده بود و انتهای منحنی بقا فقط شامل مورد سانسور شده است بنابراین، پس از برازش مدل های شفایافته پارامتریک آمیخته و ناآمیخته بیزی وایبل، لگ لجستیک و داگم نوع اول و سنجش نیکویی برازش آنها با شاخص Deviance Information Criteria(DIC) ، برای مقایسه عملکرد پیش بینی مدل های فوق با مدل شفایافته پارامتریک بیزی از شاخص RMSE استفاده شد.

برای انجام تحلیل بیزی مدل های شفایافته آمیخته و ناآمیخته، از توزیع پیشین نرمال برای بردار ضرایب گرسیونی استفاده شد همچنین، از توزیع پیشین یکنواخت برای تولید پارامترهای شکل توزیع های وایبل، لگ لجستیک و داگم استفاده گردید. توزیع پسین توام پارامترهای مدل از ترکیب توزیع پیشین توام با تابع درستنمایی حاصل شد و با روش نمونه گیری گیبز از توزیع پسین توام ۱۰۱۰۰۰۰نمونه برای هر پارامتر تولید شد. ۱۰۰۰۰ نمونه شبیه سازی شده اول به عنوان burn-in کنار گذاشته شد و برای برخورداری از نمونه های تقریبا ناهمبسته، نمونه ها با فاصله ۵۰ انتخاب شدند و در نهایت خلاصه های پسین مورد نظر براساس ۲۰۰۰۰ نمونه بدست آمد. برآوردهای بیزی پارامترها همان میانگین نمونه های گیبز استخراج شده از توزیع پسین توام می باشد. همگرایی الگوریتم مونت کارلوی زنجیر مارکوفی با استفاده از نمودارهای پیشینه و خود همبستگی برای نمونه های شبیه سازی شده مانیتور شد.

در نهایت، برای مقایسه عملکرد پیش بینی مدل ها، ۵۰۰ مجموعه داده به حجم ۲۵۲ برای چهار سناریو با مقادیر مختلف EPV (۲،۳، ۵ و ۷) شبیه سازی شد. زمان های سانسورینگ از توزیع نمایی تولید شد. زمان های بقا از مدل کاکس شبیه سازی شد و از توزیع نمایی برای تولید خطرات پایه در مدل کاکس استفاده شد.

متغیرهای پیش گو مستقل از یکدیگر در نظر گرفته شد و همه متغیرهای پیوسته از توزیع نرمال و همه متغیرهای طبقه ای از توزیع دوجمله ای شبیه سازی شدند. در تمامی سناریوهای شبیه سازی، همانند داده واقعی، داده ها به گونه ای تولید شد که از زمان ۲۵ تا زمان ۵۲ (پایان مطالعه) هیچ رویدادی رخ ندهد تا شرایط استفاده از مدل های شفایافته فراهم گردد. سرانجام، عملکرد پیش بینی تمامی مدل ها به ازای مقادیر مختلف EPV با شاخص RMSE مورد ارزیابی قرار گرفت.

یافته‌ها: براساس یافته ها ، بهترین عملکرد پیش بینی در داده واقعی و نیز داده های شبیه سازی با مقادیر مختلف EPV به ترتیب مربوط به مدل های شفایافته پارامتریک بیزی و جنگل بقای تصادفی و مدل های کاکس تحت روش های مجازات شده است.

براساس مقادیر C-index و Brier Score در ۵۰۰ مجموعه داده آموزشی و آزمودنی، عملکرد درونی و بیرونی مدل جنگل بقای تصادفی نزدیک به هم است ($C\text{-index}_{train}$=۰/۶۱۸, $C\text{-index}_{test}$= ۰/۶۵۳) و ($Brier\ Score_{train}$=۰/۰۱۷, $Brier\ Score_{test}$= ۰/۰۲۶) و در تمامی سناریوهای شبیه سازی با افزایش مقدار EPV همواره در عملکرد پیش بینی مدل جنگل بقای تصادفی پیشرفت دیده شد یعنی برای EPV=۲ مقدار شاخص Root Mean Square Error (RMSE) برابر ۰/۱۴۲ و برای EPV=۷، RMSE =۰/۰۸۸ بود.

براساس میانگین معیار Brier Score در ۵۰۰ داده آزمودنی دقت پیش بینی هر سه روش مجازات شده یکسان است ($Brier\ Score_{test}$=۰/۰۲۸) ولی براساس C-index دقت پیش بینی ریج، بیش از دو روش لاسو و الاستیک نت است زیرا بزرگترین میانگینC-index در ۵۰۰ داده آزمودنی را دارا بود($C\text{-index}_{test}$=۰/۸۱۰) همچنین در تمامی سناریوهای شبیه سازی شده ریج برترین عملکرد پیش بینی را دارا است و الاستیک نت در جایگاه بعدی قرار دارد.

در مورد مدل های شفایافته بیزی، براساس شاخص DIC مدل شفایافته ناآمیخته داگم بیزی هم در داده واقعی (DIC=۱۱۰/۷) و هم داده های شبیه سازی شده بهترین برازش را نسبت به سایر مدل های شفایافته دارا است اما عملکرد پیش بینی مدل های شفایافته براساس شاخصRMSE در داده واقعی بسیار نزدیک به هم دیده شد، بهترین عملکرد پیش بینی متعلق به مدل شفایافته ناآمیخته داگم بیزی بود (RMSE=۰/۰۱۷)

و بدترین عملکرد پیش بینی را مدل شفایافته وایبل داشت (RMSE=۰/۰۲۷). همچنین در تمامی سناریوهای شبیه سازی با مقادیر مختلف EPV مدل شفایافته ناآمیخته داگم بیزی برترین عملکرد پیش بینی را دارا بود.

بحث و نتیجه‌گیری: یافته های مطالعه حا ضر ن شان داد که مدل های شفایافته پارامتریک بیزی بهترین انتخاب برای حل مسئلهEPVکم ا ست در صورتی که تمامی رویدادها در ابتدای مطالعه رخ داده با شند و هیچ رویدادی در انتهای منحنی کاپلان مایر ظاهر نشده با شد. مدل های کاکس تحت روش های مجازات شده نیز پس از جنگل بقای تصادفی در رفع مشکل EPVکم می توانندکمک کننده باشند. در مجموع هرچند که در دقت پیش بینی مدل های مورد پژوهش تفاوت دیده شد اما برتری بین مدل ها خیلی چشم گیر و فاحش نیست. بنابراین شاید بتوان گفت علاوه بر مدل های شفایافته پارامتریک بیزی، عملکرد پیش بینی روش ناپارامتری جنگل بقای تصادفی و نیمه پارامتری کاکس تحت روش های مجازات شده قابل قبول است. پس اگر محقق تمایلی به انجام ا ستنباط بیزی ندا شته با شد، جنگل بقای ت صادفی و روش های مجازات شده می توانند مدل های منتخب برای حل پدیده EPVکم در داده های با بعد کم باشند.

کلید واژه‌ها: جنگل بقا تصادفی، داده های با بعد کم، رویداد کم، رگرسیون مجازات شده، مدل شفا یافته بیزی

Abstract

Background and aim : Nowadays, with significant advances in all areas of medical science, a significant proportion of patients with various types of diseases have long-term survival. So in many cases we have survival data that the number of patients experiencing death from the disease is much lower than that of censored cases. On the other hand, as we have many independent

variables in many studies, therefore, the ratio of the number of events to the number of independent variables or better expression of Event per Variable (EPV) value should be considered. Based on simulation studies, EPV between 10 and 20 is recommended. When EPV is low, the use of classic models such as Cox is not recommended because the calculated coefficients are not reliable.

In high-dimensional data, the data that the number of predictors is much larger than the number of observations, EPV is always low. Therefore, if you are encountered such the data, without checking the EPV value, the methods such as penalized methods will greatly reduce the variance of classical models directly. But in the case of low-dimensional data, data that the number of variables is less than the number of observations, either the amount of EPV is acceptable so the use of classic models is unrestricted or EPV is low as a result the use of the classical model will be problematic for this data so, in order to overcome the over-fitted classical model, one solution must be found to improve the model variance.

Because of the importance of the EPV index and since few studies have evaluated low dimensional and low EPV data, this study has dealt with it. And given the nature of the data used, in particular, this study aims to compare the performance of the Random Survival Forest Model, the penalized Cox Models, and the Bayesian cure Model as three different nonparametric, semi-parametric, and parametric approaches in low-dimensional data with low EPV is done.

Method: In this study 252 dialysis patients were studied. Due to 35 death events and 19 variables, the EPV value was approximately 2. The data were randomly divided into two test and train sets, and this procedure was repeated 500 times. The random survival forest models and Cox models based on penalized methods (Lasso, Ridge and Elastic Net), as the first and second proposed approaches in low dimensional data and low EPV, were implemented for training data,

respectively and evaluation of the prediction accuracy of these models in the test data was provided using calibration slope (CS), C-index and Brier Score (BS) indices.

In addition, since the data in the study, all events occurred at the beginning of the Kaplan Meier(K-M) curve and the end of the survival curve only included the censored case, therefore, After fitting the Weibull, Log-Logistic and dagum Bayesian cure Models, the RMSE index was used to compare the predictive performance of the above models with the Bayesian Parametric cure Model.

For performing Bayesian analysis of mixture and non-mixture cure models, from the prior normal distribution was used for the vector of regression coefficients. Also, the prior uniform distribution was used to generate the shape parameters of the Weibull, Log-Logistic and Dagum distributions. Also, the joint posterior distribution of the model parameters obtained with the combination of the joint prior distribution with the likelihood function and with the Gibbs sampling method, 101,000,0 samples were produced for each parameter. 10,000 first simulated samples were discarded as burn-in, and to obtain almost uncorrelated samples, the samples were selected with lag 50. Finally, the posterior summaries were obtained based on 20,000 samples. The Bayesian estimates of the parameters are the average of Gibbs samples extracted from the joint posterior distribution. Convergence of the MCMC algorithm was monitored by history and autocorrelation plots for the simulated samples. Inferences were obtained using OpenBUGS and R Softwares.

In addition, to compare the prediction performance of the models, 500 data sets with the size of 252 were simulated for four scenarios with different EPV values (2, 3, 5 and 7). Censoring times were generated by exponential distribution. Survival times were generated regarding Cox proportional hazard model and the exponential distribution applied to generate baseline hazard

in cox model. The predictors were independent of each other and all continuous predictors were generated from normal distribution and all categorical variables were produced from binomial distribution. In all simulated scenarios, the data were generated such that no event occurred from the time 25 until the end of the study to provide conditions for use of the cure models. Finally, the predictive performance of all models was evaluated for different EPV values.

Findings: Based on the results, the best prediction performance in the real data, as well as the simulation data with different EPV values, are related to the Bayesian parametric cure models and random survival forest and Cox models with penalized methods, respectively.

Based on the values of the C-index and Brier Score in 500 datasets of train and test, the internal and external performance of the Random Survival Forest model are close to each other ($C\text{-index}_{test}=0.653$, $C\text{-index}_{train}=0.618$) and ($\text{Brier Score}_{test}= 0.026$, $\text{Brier Score}_{train}=0.017$) and in all the simulation scenarios, with increasing EPV value, there was always an improvement in the predictive performance of the Random Survival Forest model, that is, for EPV=2 the value of Root Mean Square Error (RMSE) index was 0.142 and for EPV=7, RMSE = 0.088.

Based on the mean of Brier Score in 500 testing sets, the predictive accuracy of all three penalized methods is the same ($\text{Brier Score}_{test}=0.028$) but based on C-index, the predictive accuracy of Ridge is more than Lasso and Elastic Net because it had the highest mean of C-index in 500 testing sets ($C\text{-index}_{test}=0.810$) also, in all simulated scenarios, Ridge had the best predictive performance and Elastic Net is in the next place.

In the case of Bayesian cure models, based on the Deviance Information Criteria (DIC) index, the Bayesian Dagum non-mixture cure model has the best fit in both real data (DIC =110.7) and simulated data compared to other cure models but the predictive performance of the cure models based on the RMSE index was very close to each other in real data, the best predictive

performance belonged to the Bayesian non-mixure cure model with Dagum distribution (RMSE = 0.017) and the Bayesian Weibull cure models had the worst prediction accuracy (RMSE = 0.027). Also in all simulation scenarios with different EPV values, for all EPV values, the Bayesian non-mixure cure model with Dagum distribution has the best predictive performance.

Conclusion: The findings of the present study showed that Bayesian parametric cure models are the best choice for solving the EPV problem if all events occurred at the beginning of the study and no events appeared at the end of the Kaplan-Meier curve. Cox models based on penalized methods can also help reduce EPV problem after random forest survival. In general, although there was a difference in the predictive accuracy of the studied models, the superiority between the models is not very significant. So perhaps we can say that performance of all three models of nonparametric random survival forest, semi-parametric Cox methods based on penalized methods and parametric Bayesian cure are acceptable. Therefore, the researcher can then select any of the above models to solve the low EPV in low dimensional data.

Key words: Random Survival Forest, low-dimensional data, few event, Penalized Regression, Bayesian Cure Model

Kerman University of Medical Sciences

Faculty of Health

In Partial Fulfillment of the requirements for the Degree Ph.D in Biostatistics

Title :

Applying the Random Survival Forest Model in low-dimensional data with few events and compare it with Cox Regression analysis under Penalized Regression methods, and Bayesian Parametric Cure Model

By :

Shideh Rafati

Supervisor :

Dr. Abbas Bahrampour | Dr. Mohammad Reza Baneshi

Thesis No : Year : 1398-1399

١. Mogensen UB, Ishwaran H, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. Journal of statistical software. 2012;50(11):1.
٢. Kleinbaum DG, Klein M. Survival analysis: Springer; 2010.
٣. Baneshi MR, Bahmanbijari B ,Mahdian R, Haji-Maghsoodi S. Comparison of Nasal Intermittent Positive Pressure Ventilation and Nasal Continuous Positive Airway Pressure Treatments Using Parametric Survival Models. Iranian journal of pediatrics. 2014;24(2):207.
٤. Baneshi MR, Farkhani EM, Haji-Maghsoudi S. Assessment of the importance of a new risk factor in prediction models. Iranian Red Crescent Medical Journal. 2016;18(2).
٥. Nikbakht R, Bahrampour A. Determining factors influencing survival of breast cancer by fuzzy logistic regression model. Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences. 2017;22.
٦. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. Journal of clinical epidemiology. 2016;76:175-٨٢.
٧. Ojeda FM, Müller C, Börnigen D, Tregouet D-A, Schillert A, Heinig M, et al. Comparison of Cox model methods in a low-dimensional setting with few events. Genomics, proteomics & bioinformatics. 2016;14(4):235-43.
٨. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. Journal of clinical epidemiology. 1995;48(12):1503-10.
٩. Verweij PJ, Van Houwelingen HC. Penalized likelihood in Cox regression. Statistics in medicine. 1994;13(23-24):2427-36.
١٠. Tibshirani R. The lasso method for variable selection in the Cox model. Statistics in medicine. 1997;16(4):385-95.
١١. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2005;67(2):301-20.
١٢. Tibshirani RJ, Taylor J. Degrees of freedom in lasso problems. The Annals of Statistics. ٢٠١٢;٤٠(٢):١١٩٨-٢٣٢.
١٣. Rafati S, Baneshi MR, Hassani L, Bahrampour A. Comparison of penalized cox regression methods in low-dimensional data with few-events: An application to dialysis patients' data. Journal of research in health sciences. 2019;19(3.(
١٤. Yosefian I, Mosa Farkhani E, Baneshi MR. Application of random forest survival models to increase generalizability of decision trees: a case study in acute myocardial infarction. Computational and mathematical methods in medicine. 2015.
١٥. Radespiel-Tröger M, Rabenstein T, Schneider HT, Lausen B. Comparison of tree-based methods for prognostic stratification of survival data. Artificial Intelligence in Medicine. ٢٠٠٣;٢٨(٣):٣٢٣-٤١.
١٦. Banerjee M, George J, Song EY, Roy A, Hryniuk W. Tree-based model for breast cancer prognostication. Journal of clinical oncology. 2004;22(13):2567-75.
١٧. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees: CRC press; 1984.
١٨. Gordon L, Olshen RA. Tree-structured survival analysis. Cancer treatment reports. ١٩٨٥;٦٩(١٠):١٠٦٥-٩.

١٩. LeBlanc M, Crowley J. Relative risk trees for censored survival data. Biometrics. ١٩٩٢:٤١١-٢٥.
٢٠. LeBlanc M, Crowley J. Survival trees by goodness of split. Journal of the American Statistical Association. 1993;88(422):457-67.
٢١. Segal MR. Regression trees for censored data. Biometrics. 1988:35-47.
٢٢. Breiman L. Random forests. Machine learning. 2001;45(1):5-32.
٢٣. Seni G, Elder JF. Ensemble methods in data mining: improving accuracy through combining predictions. Synthesis Lectures on Data Mining and Knowledge Discovery. ٢٠١٠;٢(١):١-١٢٦.
٢٤. Breiman L. Bagging predictors. Machine learning. 1996;24(2):123-40.
٢٥. Rostami M, Garrusi B, Baneshi MR. A study on the use of bootstrap aggregation methods in estimation of stable parameters. Journal of Biostatistics and Epidemiology. 2016;2(2):104-10.
٢٦. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. The annals of applied statistics. 2008:841-60.
٢٧. Genuer R, Poggi J-M ,Tuleau C. Random Forests: some methodological insights. arXiv preprint arXiv:08113619. 2008.
٢٨. Noori S, Nourijelyani K, Mohammad K, Niknam M, Mahmoudi M, Andonian L, et al. Random Forests Analysis: a Modern Statistical Method for Screening in High-Dimensional Studies and its Application in a Population-Based Genetic Association Study. 2011.
٢٩. Hothorn T, Bühlmann P, Dudoit S, Molinaro A, Van Der Laan MJ. Survival ensembles. Biostatistics. 2005;7(3):355-73.
٣٠. Kim S, Chen M-H, Dey DK, Gamerman D. Bayesian dynamic models for survival data with a cure fraction. Lifetime Data Analysis. 2007;13(1):17-35.
٣١. Ali Akbari Khoei R, Bakhshi E, Azarkeivan A, Biglarian A. Application of cure models in survival analysis of thalassemia major disease. Razi Journal of Medical Sciences. ٢٠١٥;٢٢(١٣١):٧١-٩.
٣٢. Rahimzadeh Kiwi M, Hajizadeh E, Feyzi S. Assessment of factor effectiveness on the bilateral corneal graft rejection in the keratoconus with cure frailty model. Research in Medicine. ٢٠١٠;٣٤(٢):١١٧-٢٢.
٣٣. Corbière F, Joly P. A SAS macro for parametric and semiparametric mixture cure models. Computer methods and programs in biomedicine. 2007;85(2):173-80.
٣٤. Hoseini M, Bahrampour A, Mirzaee M. Comparison of weibull and lognormal cure models with cox in the survival analysis of breast cancer patients in Rafsanjan. Journal of research in health sciences. 2017;17(1.(
٣٥. Datema FR, Moya A, Krause P, Bäck T, Willmes L, Langeveld T, et al. Novel head and neck cancer survival analysis approach: random survival forests versus Cox proportional hazards regression. Head & neck. 2012;34(1):50-8.
٣٦. Laas E, Hamy A-S, Michel A-S, Panchbhaya N, Faron M, Lam T, et al. Impact of time to local recurrence on the occurrence of metastasis in breast cancer patients treated with neoadjuvant chemotherapy: A random forest survival approach. PloS one. 2019;14(1):e0208807.
٣٧. Hamidi O, Poorolajal J, Farhadian M, Tapak L. Identifying important risk factors for survival in kidney graft failure patients using random survival forests. Iranian journal of public health. 2016;45(1):27.

٣٨. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. Circulation: Cardiovascular Quality and Outcomes. 2011;4(1):39-45.
٣٩. Friedel G, Fritz P, Goletz S, Kristen R, Brinkmann F, Dierkesmann R, et al. Postoperative survival of lung cancer patients: are there predictors beyond TNM? Anticancer research. ٢٠١٣;٣٣(٤):١٦٠٩-١٩.
٤٠. Porzelius C, Schumacher M, Binder H. Sparse regression techniques in low-dimensional survival data settings. Statistics and Computing. 2010;20(2):151-63.
٤١. Benner A, Zucknick M, Hielscher T, Ittrich C, Mansmann U. High-dimensional Cox models: the choice of penalty as part of the model building process. Biometrical Journal. ٢٠١٠;٥٢(١):٥٠-٦٩.
٤٢. Ambler G, Seaman S, Omar R. An evaluation of penalised survival methods for developing prognostic models with rare events. Statistics in medicine. 2012;31(11-12):1150-61.
٤٣. Jafari-Koshki T, Mansourian M, Mokarian F. Exploring factors related to metastasis free survival in breast cancer patients using Bayesian cure models. Asian Pac J Cancer Prev. ٢٠١٤;١٥(٢٢):٩٦٧٣-٨.
٤٤. Baghestani AR, Moghaddam SS, Majd HA, Akbari ME ,Nafissi N, Gohari K. Application of a non-mixture cure rate model for analyzing survival of patients with breast cancer. Asian Pac J Cancer Prev. 2015;16(16):7359-63.
٤٥. Achcar JA, Coelho-Barros EA, Mazucheli J. Cure fraction models using mixture and non-mixture models. Tatra Mountains Mathematical Publications. 2012;51(1):1-9.
٤٦. Santos MRd, Achcar JA, Martinez EZ. Bayesian and maximum likelihood inference for the defective Gompertz cure rate model with covariates: an application to the cervical carcinoma study. Ciência e Natura. 2017;39(2):244-58.
٤٧. Coelho-Barros EA, Achcar JA, Mazucheli J. Cure Rate Models Considering The Burr XII Distribution in Presence of Covariate. Journal of Statistical Theory and Applications. ٢٠١٧;١٦(٢):١٥٠-٦٤.
٤٨. Martinez EZ, Achcar JA. A new straightforward defective distribution for survival analysis in the presence of a cure fraction. Journal of Statistical Theory and Practice. 2018;12(4):688-703.
٤٩. Swain PK, Grover G, Goel K. Mixture and Non-Mixture Cure Fraction Models Based on Generalized Gompertz Distribution under Bayesian Approach. Tatra Mountains Mathematical Publications. 2016;66(1):121-35.
٥٠. Ibrahim NA, Taweab F, Arasan J. A parametric non-mixture cure survival model with censored data. Computational Problems in Engineering: Springer; 2014. p. 231-8.
٥١. Naseri P, Baghestani AR, Momenyan N, Akbari ME. Application of a mixture cure fraction model based on the generalized modified weibull distribution for analyzing survival of patients with breast cancer. International Journal of Cancer Management. 2018;11(5.(
٥٢. Martinez EZ, Achcar JA, Jácome AA, Santos JS. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. Computer methods and programs in biomedicine. 2013;112(3):343-55.
٥٣. Dietrich S. Investigation of the machine learning method Random Survival Forest as an exploratory analysis tool for the identification of variables associated with disease risks in complex survival data. 2016.

بسمه تعالی

تاریخ .....

شماره .....

پیوست .....

صورتجلسه دفاع از پایان نامه

دانشگاه علوم پزشکی کرمان

مدیریت تحصیلات تکمیلی دانشگاه

جلسه دفاعیه پایان نامه تحصیلی خانم شیده یقینی دانشجوی دکتری تخصصی (Ph.D) رشته آمار زیستی دانشکده بهداشت دانشگاه علوم پزشکی کرمان تحت عنوان " به کار گیری مدل جنگل بقای تصادفی در داده های با بعد کم و رویداد کم و مقایسه آن با تحلیل رگرسیون کاکس تحت روش های رگرسیونی جهازات شده و سن شفا یافته پارامتریک بیزی "

درساعت ۱۷ روز سه شنبه مورخ ۹۹/۳/۱۲ با حضور اعضای محترم هیات داوران به شرح ذیل:

سمت	نام و نام خانوادگی	امضا
(استاد/ان) راهنما	آقای دکتر عباس بهرام پور آقای دکتر محمدرضا باقری	
(استاد/ان) مشاور		
ج. عضو هیات داوران (داخلی)	آقای دکتر یونس جهانی	
چ. عضو هیات داوران (داخلی)	خانم دکتر مقدمه میرزایی	
د. عضو هیات داوران (خارجی)	آقای دکتر حبیب الله اسماعیلی	
د. عضو هیات داوران (خارجی)	آقای دکتر آرات فیضی	
ه. نماینده تحصیلات تکمیلی	آقای دکتر عباس بهرام پور	

تشکیل گردید و ضمن ارزیابی به شرح پیوست با درجه عالی و نمره ۱۹,۸۲ مورد تایید قرار گرفت.

مهر و امضاء معاون آموزشی

۹۹/۳/۱۲