Back to previous page

ProQuest

Can large earthquakes be predicted by the evaluation of accelerated moment release?

Dunham, John. California State University, Long Beach, ProQuest Dissertations Publishing, 2009. 1472315.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Dunham%2C+John&isbn=9781109472936&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Accelerated moment release (AMR) is the name given to the theory that there is an acceleration of the release of seismic energy preceding large earthquakes (>=M6.5). It is postulated that this acceleration can be used to predict the occurrence of large earthquakes. The goal of this project is to determine if the theory of accelerated moment release is supported by the data and analysis by sound, objective statistical methods.

The primary approach is to test the model of King and Bowman using observed seismicity in northern, central, and southern California. The model of King and Bowman is evaluated using catalogs of observed data. The validity of the model is determined by comparing a regression of a power-law time-to-failure regression curve representing AMR to that of a linear regression of the same data representing a steady release of seismic energy.

Details

Advisor	Suaray, Kagba
Author	Dunham, John
Identifier / keyword	Applied sciences; Pure sciences; Earth sciences
Number of pages	70
ProQuest document ID	305179611
Publication title	ProQuest Dissertations and Theses
Publication year	2009

A comparison of Continuous Turbulence and Statistical Discrete Gust for the development of aircraft design loads

Byrne, Shaun. California State University, Long Beach, ProQuest Dissertations Publishing, 2009. 1481582.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Byrne%2C+Shaun&isbn=9781109657753&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Atmospheric turbulence can cause internal forces that design large portions of aircraft structure. The current gust loads criterion, Continuous Turbulence, is applied to the structure as velocity Power Spectral Density (PSD). Gust velocity is represented as stationary, random, and Gaussian. Structural response is obtained through probabilistic frequency domain analysis. Statistical Discrete Gust (SDG) is a deterministic time domain analysis providing tuned gust histories representative of large gusts that design primary structure. SDG assumes gust velocity has a stretched exponential distribution, which is more representative of observed data. This paper presents a method of deriving load cases using the gust profiles of SDG scaled to meet the continuous turbulence envelope. Statistical assumptions of the current criterion and SDG are presented. The implication to developed design loads is discussed. A structural response simulation is used to illustrate the design loads development and their statistically appropriate nature compared to the continuous turbulence criterion.

Details

Advisor	Suaray, Kagba
Author	Byrne, Shaun
Identifier / keyword	Applied sciences; Pure sciences
Number of pages	36
ProQuest document ID	305179850
Publication title	ProQuest Dissertations and Theses
Publication year	2009

Repeated measures modeling in the business environment

Grogan, Tristan. California State University, Long Beach, ProQuest Dissertations Publishing, 2011. 1493124.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Grogan%2C+Tristan&isbn=9781124622859&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Several marketing applications of statistics require repeated measures modeling techniques. For example, to quantify the effectiveness of a new advertising campaign, stores may be measured repeatedly over time. Failure to correctly specify a repeated measures design can lead to artificially low p-values and incorrect conclusions.

This project will analyze several different repeated measure designs using real data from a consumer packaged goods (CPG) beverage company. Three business questions will be proposed and answered by this paper. Linear mixed models and generalized mixed models will be utilized.

Details

Advisor	Suaray, Kagba
Author	Grogan, Tristan
Identifier / keyword	Pure sciences
Number of pages	35
ProQuest document ID	866321058
Publication title	ProQuest Dissertations and Theses
Publication year	2011

The cummulative incidence estimate versus the Kaplan Meier method in medical research

Dang, Ha. California State University, Long Beach, ProQuest Dissertations Publishing, 2010. 1486385.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Dang%2C+Ha&isbn=9781124247755&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

The Kaplan Meier (KM) method is very popular in medical research, but many researchers are not aware of its deficiencies when handling studies that have multiple failure types and/or competing risks. The Cumulative Incidence (CI) estimate, has been around for some time yet is rarely used, would be an excellent choice to implement when competing risks occur. Unlike the KM method, which only estimates the survival probability based on the non-censoring group, the CI estimates each failure type separately. To illustrate the superior performance of CI estimate in theory and practice, the CI estimate and the KM method will be compared and contrasted in simulations and with clinical solid tumor data. Baseline progression free survival is accurately estimated for use in further analysis at the Children's Oncology Group.

Details

Advisor	Suaray, Kagba
Author	Dang, Ha
Identifier / keyword	Applied sciences; Pure sciences
Number of pages	79
ProQuest document ID	757004704
Publication title	ProQuest Dissertations and Theses
Publication year	2010

Subgroup analysis based on prognostic and predictive gene signatures for adjuvant chemotherapy in early-stage non-small-cell lung cancer patients

Pluta, Dustin. California State University, Long Beach, ProQuest Dissertations Publishing, 2015. 1589644.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac stage+non-small-cell+lung+cancer+patients&issn=&date=2015-01-01&volume=&issue=&spage=&au=Pluta%2C+Dustin&isbn=9781321776577&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

In treating patients diagnosed with Stage I non-small-cell lung cancer, doctors must choose between surgery and Adjuvant Cisplatin-Based Chemotherapy (ACT). For patients with resected stages IB to IIIA, clinical trials have shown a survival advantage from 4-15% with the adoption of ACT. However, due to the inherent toxicity of chemotherapy, it is necessary for doctors to identify patients whose chance of success with ACT is sufficient to justify the risks. This project seeks to use gene expression profiling in the development of a statistical decision-making algorithm to identify patients whose survival rates will improve from ACT treatment. Using data from the National Cancer Institute, the Cox-Proportional-Hazards regression model will be used to determine a feasible number of genes that are strongly associated with the treatment-related patient survival. Considering treatment groups separately, patients are assigned a risk category determined by survival time. These risk categories are used to develop a random forest classification model to identify patients who are likely to benefit from chemotherapy treatment. The probability of significant benefit from chemotherapy is then predicted using a regression survival tree. This model allows the prediction of a new patient's prognosis and the likelihood of survival benefit from ACT treatment based on a small number of gene expression levels.

Details

Advisor	Moon, Hojin
Author	Pluta, Dustin
Identifier / keyword	Pure sciences; Biological sciences; Gene signature; Lung cancer; Random forest
Number of pages	81
ProQuest document ID	1688738932
Publication title	ProQuest Dissertations and Theses
Publication year	2015

Predictive models of cytotoxicity as mediated by exposure to environmental toxicants and drugs

Cong, Ming. California State University, Long Beach, ProQuest Dissertations Publishing, 2015. 1602562.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Cong%2C+Ming&isbn=9781339177755&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Predicting cytotoxicity is a challenging task because of the complex biological mechanisms behind it. Cytotoxicity due to toxin, biologically produced poison, is known to play a substantial role in disease process. Two objectives in this research are to derive robust general predictive cytotoxicity models to minimize unnecessary toxicity. The first objective is to build novel predictive statistical models for cytotoxicity data based on lymphoblastoid cell lines obtained from *in vitro* studies. This might be an important step for accomplishing a goal in biomedical/biophamarceutical research, by obtaining the best medical outcomes by minimizing toxicity in regard to a person's genomic profile. The second objective is to build predictive models to predict population-level cytotoxicity for unknown compounds based on chemical structural features. Since environmental chemical compounds has greatly influence on human health, the predictive statistical models built within this objective could be helpful to government agencies for the relevant decision-making.

Details

Advisor	Moon, Hojin
Author	Cong, Ming
Identifier / keyword	Pure sciences; Biological sciences; Environmental toxicants
Number of pages	48
ProQuest document ID	1734864292
Publication title	ProQuest Dissertations and Theses
Publication year	2015

Model uncertainty and model averaging in the estimation of benchmark dose

Kim, Steven B.. California State University, Long Beach, ProQuest Dissertations Publishing, 2010. 1486380.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Kim%2C+Steven+B.&isbn=9781124247700&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Food-borne infection is caused by intake of contaminated foods or beverages. Since there is no "default model" for microbial risk assessment, model averaging in the estimation of benchmark dose has been studied to analyze microbial dose-response experiments. In this research, several dose-response models including four two-parameter models and four three-parameter models are used. Parameters of the statistical models are estimated by maximum likelihood method. The benchmark dose is estimated by a weighted average of effective dose estimates from the eight models, and the weights are determined by Kullback information criterion to account for model uncertainty. Both model uncertainty and data uncertainty are addressed to compute the variance of the benchmark dose estimate, and a bootstrap-based 95% confidence interval of benchmark dose is constructed. To evaluate the coverage probabilities of the confidence limits, a Monte Carlo simulation study is conducted in various conditions based on a real data set in human volunteers.

Details

Advisor	Moon, Hojin
Author	Kim, Steven B.
Identifier / keyword	Applied sciences; Pure sciences
Number of pages	93
ProQuest document ID	757004572
Publication title	ProQuest Dissertations and Theses
Publication year	2010

Various considerations on performance measures for a classification of ordinal data

Nyongesa, Denis Barasa. California State University, Long Beach, ProQuest Dissertations Publishing, 2016. 10133995.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Nyongesa%2C+Denis+Barasa&isbn=9781339924977&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

The technological advancement and the escalating interest in personalized medicine has resulted in increased ordinal classification problems. The most commonly used performance metrics for evaluating the effectiveness of a multi-class ordinal classifier include; predictive accuracy, Kendall's tau-b rank correlation, and the average mean absolute error (AMAE). These metrics are beneficial in the quest to classify multi-class ordinal data, but no single performance metric incorporates the misclassification cost. Recently, distance, which finds the optimal trade-off between the predictive accuracy and the misclassification cost was proposed as a cost-sensitive performance metric for ordinal data. This thesis proposes the criteria for variable selection and methods that accounts for minimum distance and improved accuracy, thereby providing a platform for a more comprehensive and comparative analysis of multiple ordinal classifiers. The strengths of our methodology are demonstrated through real data analysis of a colon cancer data set.

Details

Advisor	Moon, Hojin
Author	Nyongesa, Denis Barasa
Identifier / keyword	Pure sciences; Biological sciences; Applied sciences; Classification trees; Colon cancer; Logistic regression; Ordinal data; Random forests; Support vector machines
Number of pages	111
ProQuest document ID	1810990473
Publication title	ProQuest Dissertations and Theses
Publication year	2016

Identification of risk factors and likelihood of benefit from adjuvant chemotherapy for early stage lung cancer patients

Chao, Ted. California State University, Long Beach, ProQuest Dissertations Publishing, 2017. 10264880.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Chao%2C+Ted&isbn=9781369773965&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

The purpose of the research is to develop a statistical decision support algorithm for patients who may benefit from Adjuvant Cisplatin/Vinorelbine (ACT) and improve their survival rates. Genome-wide microarray data will be used to identify feasible sets of genes and probe sets that constitute the gene signature. The data are available at the National Center for Biotechnology Information Gene Expression Omnibus (GSE14814). Preliminary studies have shown that high risk patients that received ACT resulted in an improved prognosis. However, low risk patients showed no benefit from ACT and the treatment was possibly detrimental to the patient. Studies using random forests models have shown that genomic markers could potentially identify a patient's risk factor and likelihood to benefit from ACT; however, it was noted that the random forests do not provide an estimate of the strength of the treatment effect, nor is it possible to clearly identify subgroups of patients with similar responses to ACT treatment. Building on this idea, Accelerated Failure Time models are used to predict the probability of benefit from receiving chemotherapy or surgery only and provide a treatment recommendation. We showed that regardless of whether the model recommended chemotherapy or surgery only, patients that followed the treatment recommendation had significantly longer survival times than patients that did not. For new patients, the model can provide the likelihood of benefit for each treatment based on a small number of genomic biomarkers.

Details

Advisor	Moon, Hojin
Author	Chao, Ted
Identifier / keyword	Pure sciences; Biological sciences
Number of pages	72
ProQuest document ID	1906299341
Publication title	ProQuest Dissertations and Theses
Publication year	2017

Nonparametric alternative to Poly-*k* test in animal tumorigenicity studies

Dill, Kristina Dionne. California State University, Long Beach, ProQuest Dissertations Publishing, 2012. 1521617.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beach+test+in%0Aanimal+tumorigenicity+studies&issn=&date=2012-01-01&volume=&issue=&spage=&au=Dill%2C+Kristina+Dionne&isbn=9781267790460&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

In the Food and Drug Administration, statistical testing for dose related trend is recommended to examine tumorigenicity in long term animal experiments. The survival-adjusted Cochran-Armitage test, known as the Poly-*k* test, is used for detecting a linear trend in the incidence of an occult tumor of interest across dose groups. Usually, in long term animal tumorigenicity studies the Poly-3 test with the shape parameter 3 is used. In this research, an age-adjusted bootstrap-based method is developed as a nonparametric alternative to the Poly-3 test in animal carcinogenicity data. An empirical distribution of the Poly-3 trend test statistic is investigated by comparing the proposed age-adjusted bootstrap method to the Poly-3 test statistic. The proposed method is applied to National Toxicology Program data sets to evaluate a dose-related trend of a test substance on the incidence of tumors. Further investigations will be conducted in a simulation study to evaluate the robustness of these tests to various Weibull-family tumor onset distributions.

Details

Advisor	Moon, Hojin
Author	Dill, Kristina Dionne
Identifier / keyword	Pure sciences; Biological sciences; Health and environmental sciences
Number of pages	89
ProQuest document ID	1223507040
Publication title	ProQuest Dissertations and Theses
Publication year	2012

Specification of sex-specific molecular biomarkers

Lopez, Karen L.. California State University, Long Beach, ProQuest Dissertations Publishing, 2012. 1517720.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac specific+molecular+biomarkers&issn=&date=2012-01-01&volume=&issue=&spage=&au=Lopez%2C+Karen+L.&isbn=9781267470812&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Ideally, treatments for patients would be standardized and uniform, based on the specific illness or disease, despite individual characteristic differences; however, numerous studies have revealed this not to be the case, since differences are seen in reactions to the same drug treatment reliant on the person's sex, thus undermining the traditional view of one-size-fits-all medicine. This study investigates the influence of sex on disease characteristics and risk factors. An algorithm is proposed to isolate a set of sex-specific genomic biomarkers using the Random Forest algorithm to rank the importance of genes from gene expression data. Cross-validation is used to isolate a feasible set of genes and to obtain performance of sex-specific biomarkers. The selected a set of sex-specific biomarkers will improve accuracy in classification of patients which will provide more effective treatment. The proposed procedure is applied to two gene expression datasets.

Details

Advisor	Moon, Hojin
Author	Lopez, Karen L.
Identifier / keyword	Pure sciences; Biological sciences; Health and environmental sciences
Number of pages	93
ProQuest document ID	1027934809
Publication title	ProQuest Dissertations and Theses
Publication year	2012

A Clinical Decision Support System for the Prevention of Genetic-Related Heart Disease

Saguilig, Lauren G.. California State University, Long Beach, ProQuest Dissertations Publishing, 2017. 10264716.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac Related+Heart+Disease&issn=&date=2017-01-01&volume=&issue=&spage=&au=Saguilig%2C+Lauren+G.&isbn=9781369805130&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Drug-induced long QT syndrome (diLQTS) is a common adverse drug reaction characterized by rapid and erratic heart beats that may instigate fainting or seizures. The onset of diLQTS can lead to torsades de points (TdP), a specific form of abnormal heart rhythm that often leads to sudden cardiac arrest and death. This study aims to understand the genetic similarities between diLQTS and TdP to develop a clinical decision support system (CDSS) to aide physicians in the prevention of TdP. Highly accurate classification algorithms, including random forests, shrunken centroid, and diagonal linear discriminant analysis are considered to build a prediction model for TdP. With a feasible set of markers, we accurately predict TdP classifications with an accuracy above 90%. The methodology used in this study can be extended to dealing with other biomedical high-dimensional data.

Details

Advisor	Moon, Hojin
Author	Saguilig, Lauren G.
Identifier / keyword	Pure sciences
Number of pages	42
ProQuest document ID	1908528209
Publication title	ProQuest Dissertations and Theses
Publication year	2017

A statistical method for selection, classification, and network construction in genetic systems

Apitz, Juan Carlos. California State University, Long Beach, ProQuest Dissertations Publishing, 2016. 10127023.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Apitz%2C+Juan+Carlos&isbn=9781339850610&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

From the combination of Mendelian Genetics and Biometrics in the early 1900s to the completion of the Human Genome Project in 2003, statistical analysis has played a key role in the advancement of genetics and genomic research. Although much progress has been achieved in these fields, singular cures to genetic disease such as cancer still elude us.

The purpose of this work is to formulate a statistical framework to aid the understand- ing of genetic expression as a complex system. The primary idea is that the complexity of genomic systems can be represented as a network and we use statistical methods to infer the network structure. As a case study, we employ a regularization method called the Elastic Net (a combination of LASSO and Ridge regression) on microarray breast cancer data and identify an informative gene subset. Once the informative gene subset is identified, we infer its network structure.

This methodology proves to be effective in several ways. First, Elastic Net regularization allows us to solve problems such as $y = X\beta + E$, where the design matrix X has N rows, p columns, and is rank-deficient by virtue of p being much greater than N. A design matrix like X is a typical feature of microarray data which motivates the application of this method in the genomic setting. The solution that results is sparse which is effectively a method for variable (informative gene) selection. Second, we show, that once the informative gene set is identified, we can use it as an input to a simple logistic regression model and perform cancer type classification. The resulting classification accuracy rate is comparable to that of more complex classification models. Finally, we use linear regression regularized via the Elastic Net to construct a co-expression network based on the estimated regression coefficients.

Details

Advisor	Moon, Hojin
Author	Apitz, Juan Carlos
Identifier / keyword	Pure sciences; Biological sciences
Number of pages	48
ProQuest document ID	1808950089
Publication title	ProQuest Dissertations and Theses
Publication year	2016

Weighted voting ensembles for high dimensional data

Hordoan, Liliana A.. California State University, Long Beach, ProQuest Dissertations Publishing, 2015. 1584935.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Hordoan%2C+Liliana+A.&isbn=9781321603477&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

In recent years, the understanding and development of microarray data has grown intensively to the benefit of medical science. One of these areas involves the application of statistical algorithms to categorize disease, treatments, cancers, outcomes, etc., especially for high-dimensional data. This thesis investigates the application of two different types of decision voting schemes for the classification of outcomes using microarray data. A weighted adjusted voting scheme is compared to the standard majority voting scheme for classification algorithms in different type of ensemble models. The investigation starts with decision trees as base classifiers, and then works on the improvement of the ensemble structure to investigate how the weighted adjusted voting scheme performs on actual microarray data. Due to the structure of high-dimensional data, cross-validation is used to evaluate the validity of the statistical analysis. Variable importance is considered in this research to improve model efficiency by selecting top-ranked genes via Random Forest. Then accuracy is assessed on different ensemble methods to draw conclusions on the performance of weighted voting scheme compared to average majority voting.

Details

Advisor	Moon, Hojin
Author	Hordoan, Liliana A.
Identifier / keyword	Pure sciences
Number of pages	67
ProQuest document ID	1660972600
Publication title	ProQuest Dissertations and Theses
Publication year	2015

Case management system data analysis for a small medical device company

Santiago, Krystal. California State University, Long Beach, ProQuest Dissertations Publishing, 2008. 1463187.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Santiago%2C+Krystal&isbn=9781109034615&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Data from a medical device company's case scheduling system was downloaded in June 2008 for the purpose of exploring the contributing factors to a canceled case. The forecasting that is done for production volumes relies on the cases scheduled by sales representatives. It is advantageous to explore the data set to determine if there is a way to predict the cancellations.

The data set consists of 2,584 entries where the target variable of "status" is either complete or cancel. Standard data mining techniques will be used to explore and analyze the data. A subset of the data will be used to further determine the cause of the cancellation. Preliminary results show that the cancellation rate is directly related to the type of case being scheduled and how far in advance the case is scheduled.

Details

Advisor	Safer, Alan
Author	Santiago, Krystal
Identifier / keyword	Pure sciences
Number of pages	169
ProQuest document ID	304842929
Publication title	ProQuest Dissertations and Theses
Publication year	2008

Chemical measurement analysis using ANOVA

Yota, Fudjie. California State University, Long Beach, ProQuest Dissertations Publishing, 2009. 1466177.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Yota%2C+Fudjie&isbn=9781109169812&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

A critical requirement for success in today's business environment is having the highest product quality level possible. This requires some measurement activities. For some type of physical measurement systems, it is a scientific fact that no two measurements are the same. Therefore, for the measurement to be useful and valid, the measurement system must be capable to measure a characteristic with the least uncertainty (i.e., error).

This project studies the capability of a system to measure the material amount or content in a sample. In the first part of this project, Gage Repeatability and Reproducibility (GRR) using Analysis Of Variance (ANOVA) method is used to identify the significant sources of measurement variability as well as estimate the size of their effects. In the second part of the project, regression analysis is used to determine whether current the test method, "T-Method," can be replaced with an alternative faster method, "S-Method."

Details

Advisor	Safer, Alan
Author	Yota, Fudjie
Identifier / keyword	Pure sciences
Number of pages	35
ProQuest document ID	305181072
Publication title	ProQuest Dissertations and Theses
Publication year	2009

Forecasting grid capacity service prices for procurement and risk management in integra ted forward markets

Edwards, Jason Scott. California State University, Long Beach, ProQuest Dissertations Publishing, 2013. 1527482.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Edwards%2C+Jason+Scott&isbn=9781303773754&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Electric utilities are responsible for meeting customer demand in a safe and cost effective manner. Utilities produce or procure energy from generating resources to meet this demand in various intervals, from hour-ahead to twenty years or more into the future. The vast majority of the power thus procured is on a forecast basis.

Power supply and customer demand must always be perfectly balanced; otherwise damage will occur to generating units and/or the electric grid. Since forecasts of customer demand can never be perfect, there must exist sufficient generator flexibility to compensate for instantaneous fluctuations in supply or demand.

Grid capacity services provide this flexibility. Sufficiently agile generating units help ensure grid reliability by continuously matching supply and demand. Consequently, power plants capable of such flexibility have incremental value to ratepayers and grid operators. This value is realized through market awards of grid capacity services at the prevailing grid capacity service price. In this project, we use multiple regression analyses to forecast grid capacity service prices. These forecasts establish, in part, the value of generating facilities that ABC electric utility may wish to procure on behalf of its customers. Capturing the incremental value of grid capacity service helps the utility to make prudent financial decisions which affect the electric rates paid by their customers.

Details

Advisor	Safer, Alan
Author	Edwards, Jason Scott
Identifier / keyword	Pure sciences; Energy markets; GCS; Nested regression
Number of pages	38
ProQuest document ID	1512223871
Publication title	ProQuest Dissertations and Theses
Publication year	2013

Violation of assumptions in linear regression model: Regressions as remedial measures

Katan, Moshe. California State University, Long Beach, ProQuest Dissertations Publishing, 2007. 1442707.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Katan%2C+Moshe&isbn=&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

The topic of the thesis is violation of assumptions in a multivariate linear regression model and implementation of various regressions as remedies. Chapter 1 is an introductory chapter. It discusses the basics of the multivariate linear regression model and its assumptions. The subsequent chapters deal with the assumptions being violated in some form.

Chapter 2 deals with presence of outlying influential observations in a dataset. Different methods for detecting the influential observations are discussed. Several types of the remedial robust regression are studied.

Chapter 3 focuses on the violation of the assumption of constant variance of the error terms. The resulting case of heteroscedasticity is defined, and several statistical tests for its confirmation are given. The remedial weighted least-squares regression model is given a full treatment.

The regression considered in Chapter 4 deviates from the linear regression in a sense that no known relation between the response variable and the predictor variables is assumed. A number of possibilities in implementing a nonparametric regression are discussed.

Finally, Chapter 5 is devoted to the instance of multicollinearity of the predictor variables. The ridge regression is introduced and discussed at length.

Details

Advisor	Safer, Alan; Korostlera, Olga
Author	Katan, Moshe
Identifier / keyword	Pure sciences
Number of pages	82
ProQuest document ID	304708469
Publication title	ProQuest Dissertations and Theses
Publication year	2007

Statistical analysis of decision trees

Baskaya, Bengu. California State University, Long Beach, ProQuest Dissertations Publishing, 2011. 1493089.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Baskaya%2C+Bengu&isbn=9781124622446&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Decision trees have been widely used for many years in the statistical literature as powerful, effective, and easily interpretable classification algorithms that are able to automatically select relevant features. This thesis examines in detail both the tree growing phase and the underlying statistical analysis. In addition, various splitting algorithms and stopping rules are explained explicitly to obtain the most powerful tree model. This paper also utilizes numerous techniques to evaluate the accuracy of model. Moreover, the main idea of boosting and its corresponding algorithms (such as AdaBoost, AdaBoost.M1 and AdaBoost.MH) that are used to improve the accuracy of model are also discussed in this paper. Finally, we present two different applications to illustrate how decision tree and boosting method apply to live data.

Details

Advisor	Safer, Alan
Author	Baskaya, Bengu
Identifier / keyword	Pure sciences
Number of pages	96
ProQuest document ID	866321010
Publication title	ProQuest Dissertations and Theses
Publication year	2011

Kolmogoroff-Smirnoff enhancement

Barron, Dean S.. California State University, Long Beach, ProQuest Dissertations Publishing, 2007. 1451232.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac Smirnoff+enhancement&issn=&date=2007-01-01&volume=&issue=&spage=&au=Barron%2C+Dean+S.&isbn=9780549407485&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Based upon publications by Andrei Kolmogoroff with the computation $|Fn(x) - F0(x)|$, the Kolmogoroff-Smirnoff test (KS) was first introduced in 1939 by Nikolai Smirnoff in, *On the estimation of the discrepancy between empirical curves of distribution for two independent samples.* The test is classified as an omnibus nonparametric test.

However, the test considers each such ΔF_k as having equal weighting, a problematic approach. Such weaknesses are explored and a KS enhancement proposed, called oando. Based upon ranking the Δx_k and then using these ranks as weights, oando preserves the nonparametric nature of KS.

Datasets based upon published birdcall data are generated and analyzed. Perspective ratios are defined and introduced which allow assessment of results. KS and oando are compared and topics for future research explored.

Details

Advisor	Safer, Alan
Author	Barron, Dean S.
Identifier / keyword	Pure sciences
Number of pages	176
ProQuest document ID	304709782
Publication title	ProQuest Dissertations and Theses
Publication year	2007

Stochastic simulation model of the apartment building investment project

Atthayuwat, Sutthira. California State University, Long Beach, ProQuest Dissertations Publishing, 2010. 1486447.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Atthayuwat%2C+Sutthira&isbn=9781124247465&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

The traditional deterministic method for analysis of investment evaluation seems to provide incomplete information about the uncertainties of the real business climate. This paper studies the financial viability of a proposed apartment building project using the discounted cash flow (DCF) model to evaluate the single point estimates of financial performance measures, and then applies the Monte Carlo simulation model to incorporate the uncertainty of input variables into the DCF model. The results of the simulation model provide the probability distribution of the outcomes, suggesting that the central values of financial indicators are lower than those of the deterministic model. The performance measures of the projects are found to be most sensitive to the growth of the market rent, loan rate, and construction costs.

Details

Advisor	Safer, Alan
Author	Atthayuwat, Sutthira
Identifier / keyword	Applied sciences; Pure sciences
Number of pages	40
ProQuest document ID	755658522
Publication title	ProQuest Dissertations and Theses
Publication year	2010

Local spatial modeling using Geographically Weighted Regression (GWR)

Wankie, Che. California State University, Long Beach, ProQuest Dissertations Publishing, 2013. 1523087.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Wankie%2C+Che&isbn=9781303202919&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Linear regression modeling is a technique used in several fields for modeling, analyzing, and predicting the relationships between variables. In spatial analysis, regression modeling has been used extensively as a general technique to analyze geographic variables. For example, in spatial epidemiology, this technique has been used to study the spatial distribution of diseases. This thesis explored a regression technique, Geographically Weighted Regression (GWR), which unlike linear regression (ordinary least squares) accounts for spatial non-stationarity. GWR is a way of exploring spatial non-stationarity by calibrating a multiple regression model which allows different relationships to exist at different points in space. This thesis also explored the technique of GWR as a "proper" statistical model and how estimates may be obtained. Local entropy map, a statistical technique in spatial data mining, was explored to examine spatial patterns based on the bivariate relationships between the dependent and each explanatory variable. Data were analyzed to explore the performance of GWR in comparison to OLS and to examine multivariate relationships at local regions using local entropy map.

Details

Advisor	Safer, Alan
Author	Wankie, Che
Identifier / keyword	Pure sciences; Earth sciences
Number of pages	112
ProQuest document ID	1416365230
Publication title	ProQuest Dissertations and Theses
Publication year	2013

Approximating algebraic formulas for prediction bounds on regression-based estimates

Feldman, Daniel. California State University, Long Beach, ProQuest Dissertations Publishing, 2009. 1472274.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac based+estimates&issn=&date=2009-01-01&volume=&issue=&spage=&au=Feldman%2C+Daniel&isbn=9781109472110&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

When attempting to predict a single observation resulting from a new regressor variable case, it is evident that the farther from the center of the historical data base the prediction is made, the larger the uncertainty of the results coming from the regression model. This uncertainty can be expressed in terms of prediction bounds, which, in the case of a single regressor variable, appears to form hyperbolas above and below the graph of the regression line, widening as the regressor variable values move away from the center of the data base. Explicit formulas for the prediction bounds are known only when the modeling relationship has been derived using classical linear regression (i.e., "Ordinary Least Squares" or OLS), and these formulas are known to describe hyperbolas. When working with relationships derived by non-OLS regression, no explicit formulas currently exist. Bootstrap random sampling has been suggested as a method for deriving an approximation of the hyperbolic prediction bounds. It has been hypothesized that approximate algebraic formulas for these prediction bounds can be derived by fitting a rotated and translated general-form hyperbola to the bootstrap results. In addition to examining the prediction bounds, inferences about the regression parameters are made to determine their statistical significance. Thus a similar investigation into utilizing bootstrap sampling to approximate confidence bounds on the parameters is done for the non-OLS situation, as explicit formulas for these confidence bounds are already known in OLS.

Details

Advisor	Safer, Alan
Author	Feldman, Daniel
Identifier / keyword	Pure sciences
Number of pages	75
ProQuest document ID	305179634
Publication title	ProQuest Dissertations and Theses
Publication year	2009

Multiple regression analyses in the prediction of aerospace instrument costs

Tran, Linh. California State University, Long Beach, ProQuest Dissertations Publishing, 2011. 1493191.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Tran%2C+Linh&isbn=9781124622286&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

The aerospace industry has been investing for decades in ways to improve its efficiency in estimating the project life cycle cost (LCC). One of the major focuses in the LCC is the cost/prediction of aerospace instruments done during the early conceptual design phase of the project. The accuracy of early cost predictions affects the project scheduling and funding, and it is often the major cause for project cost overruns.

The prediction of instruments' cost is based on the statistical analysis of these independent variables: Mass (kg), Power (watts), Instrument Type, Technology Readiness Level (TRL), Destination: earth orbiting or planetary, Data rates (kbps), Number of bands, Number of channels, Design life (months), and Development duration (months).

This author is proposing a cost prediction approach of aerospace instruments based on these statistical analyses: Clustering Analysis, Principle Components Analysis (PCA), Bootstrap, and multiple regressions (both linear and non-linear). In the proposed approach, the Cost Estimating Relationship (CER) will be developed for the dependent variable Instrument Cost by using a combination of multiple independent variables. “The Full Model” will be developed and executed to estimate the full set of nine variables. The SAS program, Excel, Automatic Cost Estimating Integrate Tool (ACEIT) and Minitab are the tools to aid the analysis. Through the analysis, the cost drivers will be identified which will help develop an ultimate cost estimating software tool for the Instrument Cost prediction and optimization of future missions.

Details

Advisor	Safer, Alan
Author	Tran, Linh
Identifier / keyword	Applied sciences
Number of pages	78
ProQuest document ID	866340397
Publication title	ProQuest Dissertations and Theses
Publication year	2011

Modeling changes in the specific heat values of metal samples across multiple trials

Koutroulis, Mathew. California State University, Long Beach, ProQuest Dissertations Publishing, 2011. 1493140.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Koutroulis%2C+Mathew&isbn=9781124623047&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Metal samples used for the determination of specific heat values in introductory chemistry laboratory courses generally exhibit a marked change in appearance through repeated usage, suggesting that oxidation is likely occurring on the metals' surface. The presence of the oxide impurity may result in significant errors in the determination of an accurate specific heat value. To mitigate this effect, metals may be rinsed with an acid followed by deionized water before drying.

This project describes the modeling of the experimental error in the determination of the specific heat for three metals using three different treatments/rinses. Using the trial number to represent time intervals, two longitudinal models are constructed and compared. A general, three-factor linear model is presented which presents the trial number as a factor. All models presented are statistically valid representations of the data and indicate that the various factors presented do have an effect on the experimental error.

Details

Advisor	Safer, Alan
Author	Koutroulis, Mathew
Identifier / keyword	Applied sciences; Pure sciences
Number of pages	57
ProQuest document ID	866354240
Publication title	ProQuest Dissertations and Theses
Publication year	2011

Ordinal time series analysis for Air Quality Index (AQI) in San Bernardino County

Chitakasempornkul, Kessinee. California State University, Long Beach, ProQuest Dissertations Publishing, 2012. 1517602.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Chitakasempornkul%2C+Kessinee&isbn=9781267460042&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Ambient pollutant, especially ground level ozone that causes respiratory diseases, has been a great concern in Southern California. U.S. Environmental Protection Agency provides the Air Quality Index (AQI) as a tool to assist the public of health warnings. AQI for ozone is currently divided into six states depending on the level of public health concern. In statistical point of view AQI can be characterized as nonstationary ordinal-valued time series. The purpose of this study is to implement statistical models for short-term forecasting of AQI. This thesis presents a generalized linear type modeling to handle the autocorrelated ordinal time series. The model is applied with four different link functions: identity, logit, probit, and complementary log-log and their forecast performance are compared. Random time-varying covariates include past AQI state, various meteorological processes, and periodic component. Data used in this study are AQI for ozone from five monitoring stations in San Bernardino County, CA for 2004 to 2006.

For the purpose of evaluating the performance of one-day-ahead forecast, the 2007 data from the same place are used. The meteorological data are from the nearby Barstow city in San Bernardino County. The portmanteau test is used to test error autocorrelations. The partial likelihood ratio test, Akaike information criterion (AIC), and Bayesian information criterion (BIC) are used to measure the goodness of fit and compare the models. The results show the model well captures the nonstationarity in ozone process and remove the nonstationarity in residuals. Both logit and probit models correctly forecast about 85% of the observed AQI.

Details

Advisor	Kim, Sung Eun
Author	Chitakasempornkul, Kessinee
Identifier / keyword	Pure sciences; Health and environmental sciences; Earth sciences
Number of pages	130
ProQuest document ID	1033328850
Publication title	ProQuest Dissertations and Theses
Publication year	2012

A comparison of unsupervised learning techniques for detection of medical abuse in automobile claims

Yang, Li. California State University, Long Beach, ProQuest Dissertations Publishing, 2012. 1521650.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Yang%2C+Li&isbn=9781267790798&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Automobile claims abuse is a widespread problem that costs insurer and consumers billions of dollars per year in lost profits and higher premiums. Due to logistical and legal complications, however, many insurers are reluctant to classify formally abuse and fraud. Unfortunately, this removes the ability to perform supervised learning since the true classification of abuse is not known. Insurers are thus forced to employ unsupervised learning techniques to detect abusive claims.

The purpose of this project is to compare the effectiveness of three unsupervised learning methods on automobile claims medical abuse in one anonymous U.S. state. The analysis is performed on a collection of abusive behavioral patterns recommended by seasoned claims adjustors. Of the three unsupervised learning methods, two of these—K-Means and hierarchical clustering-are commonly used in multivariate statistics. The third method, PRIDIT (principal component analysis of relative to an identified distribution), is a novel technique that has the potential of not only accurately classifying abuse, but also categorizing the importance of each pattern. The merits and drawbacks of all three techniques are analyzed in this paper.

Details

Advisor	Korosteleva, Olga
Author	Yang, Li
Identifier / keyword	Pure sciences
Number of pages	35
ProQuest document ID	1223503741
Publication title	ProQuest Dissertations and Theses
Publication year	2012

Understanding software development processes through peer review data analyses

Purcell, Leitha Alicia. California State University, Long Beach, ProQuest Dissertations Publishing, 2009. 1472358.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Purcell%2C+Leitha+Alicia&isbn=9781109472721&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Background. Software development is a human-intensive activity. In the pursuit of translating product requirements into executable code, defects are introduced. There are several techniques in current software development methodologies used to find those defects, including peer reviews. This project investigates measurements of the peer review process and provides recommendations on the use of statistical techniques to predict product quality and peer review efficiency.

Methods. Base calculations of defect density and peer review efficiency are performed. Basic statistics of the calculated data are generated and individual distributions are identified. Possible methods for statistical prediction include: multivariate regression, decision trees, logistic regression, cluster analysis, neural network analysis, and statistical process control.

There are approximately 6,100 peer review records. This data are drawn from 7 projects representing over 15 years of peer reviews conducted in a company in the aerospace industry.

Details

Advisor	Korosteleva, Olga
Author	Purcell, Leitha Alicia
Identifier / keyword	Applied sciences
Number of pages	65
ProQuest document ID	305178538
Publication title	ProQuest Dissertations and Theses
Publication year	2009

Statistical capability as an element of business performance analysis

Braggs, James E.. California State University, Long Beach, ProQuest Dissertations Publishing, 2012. 1511335.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Braggs%2C+James+E.&isbn=9781267308498&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Statistical capability analysis involves superimposing the quantitative requirements of a customer onto the probabilistic outcomes of a business process. The net result is an estimate of the probability that those specific customer requirements can be met or exceeded. Armed with such information, business leaders can structure their respective processes to reduce costs, while more precisely targeting customer requirements. This article provides a case study in the use of statistical capability analysis to predict a service business' capability on performance metrics. The analysis begins by evaluating the prerequisites of statistical capability: statistical control, sufficient statistical power, and the elimination of alternative competing hypotheses. Although the use of statistical capability analysis is well known in manufacturing contexts, this article contends that the analysis is broadly applicable to most business processes, assuming the foundational assumptions of this analysis are known and the methodological requirements are met.

Details

Advisor	Korosteleva, Olga
Author	Braggs, James E.
Identifier / keyword	Applied sciences; Pure sciences
Number of pages	63
ProQuest document ID	1011321678
Publication title	ProQuest Dissertations and Theses
Publication year	2012

Bayesian monitoring of clinical trials: Examples using conjugate priors

Tromp, Mary. California State University, Long Beach, ProQuest Dissertations Publishing, 2015. 1583226.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Tromp%2C+Mary&isbn=9781321538663&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

A clinical trial can save time and resources if it incorporates Bayesian monitoring. Generally speaking, conducting Bayesian analysis is a computationally intensive task. However, in the special case of hypotheses testing for clinical trials, and, moreover, when conjugate prior distributions of parameters are used, computational complexity is reduced remarkably. This thesis presents three examples where the Bayesian monitoring is achieved with a prior density of a parameter and the likelihood function of the data belonging to conjugate families of distributions. The first example studies a heart valve trial with a Poisson rate of adverse events and a gamma prior distribution of the rate. The second example focuses on testing certain drug efficacy for lowering high blood pressure, with self-conjugate normal family of distributions. In the third example, the probability of a false positive alarm produced by a heart defibrillator is modeled with beta prior distribution conjugate to binomial likelihood function.

Details

Advisor	Korosteleva, Olga
Author	Tromp, Mary
Identifier / keyword	Pure sciences; Biological sciences; Bayesian monitoring; Clinical trials; Conjugate priors; Interim monitoring
Number of pages	26
ProQuest document ID	1655360774
Publication title	ProQuest Dissertations and Theses
Publication year	2015

Hearing loss and fluctuations in Meniere's disease/cochlear hydrops

Nguyen, Huy. California State University, Long Beach, ProQuest Dissertations Publishing, 2012. 1511372.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Nguyen%2C+Huy&isbn=9781267308863&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Meniere's disease is a clinical disorder of the inner ear defined by periodic vertigo, fluctuating hearing loss, tinnitus, and aural fullness. Cochlear Hydrops is similar to Meniere's but without vertigo. It is appropriate to mention Cochlear Hydrops because it may precede Meniere's disease.

Clinical care focuses on reducing the severity and frequency of vertigo because it is the most crippling symptom of Meniere's disease. Vertigo, tinnitus, and aural fullness are subjective; therefore many studies include hearing measurement as a secondary endpoint. Currently, there are no reliable metrics to describe hearing loss and fluctuations. The lack of information on hearing changes restricts its use as the primary endpoint in clinical trials. The purpose of this paper is to quantify both the probability of fluctuations (logistic regression) and the expected hearing losses over a period of time (longitudinal data analysis).

Details

Advisor	Korosteleva, Olga
Author	Nguyen, Huy
Identifier / keyword	Health and environmental sciences; Pure sciences
Number of pages	61
ProQuest document ID	1011321682
Publication title	ProQuest Dissertations and Theses
Publication year	2012

Generalized linear regression models for count data

Nava, Martha. California State University, Long Beach, ProQuest Dissertations Publishing, 2014. 1528005.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Nava%2C+Martha&isbn=9781303985294&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Count data are observations of only non-negative integer values (0, 1, 2, etc.). When the response variable follows a Poisson distribution, the Poisson regression may be used to model the data. The main feature of a Poisson distribution is that the mean is equal to the variance. If this condition does not hold, and the variance is much larger than the mean, overdispersion occurs. When this arises, a negative binomial regression model may be employed. If there are more zeros observed than normal for a Poisson (or negative binomial) regression, a zero-inflated Poisson (or negative binomial) regression model may be applicable. If zeros are not observed due to the data not containing zero counts, then a zero-truncated Poisson (or zero-truncated negative binomial) model may be used.

This research will explain these models with examples as well as the test for overdispersion using SAS and R software.

Details

Advisor	Korosteleva, Olga
Author	Nava, Martha
Identifier / keyword	Pure sciences; Count data; Hurdle; Over dispersion; Poisson; Zero inflated; Zero truncated
Number of pages	54
ProQuest document ID	1552915004
Publication title	ProQuest Dissertations and Theses
Publication year	2014

Statistical analysis of credit card debt collectability based on data from a debt collection agency

Yoo, Terrie. California State University, Long Beach, ProQuest Dissertations Publishing, 2016. 10131674.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Yoo%2C+Terrie&isbn=9781339894454&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

The global financial recession in 2009 brought attention to consumers and financial industries concerning the important role credit history plays regarding lending and debt repayment. Going through this financial era, especially, collection agencies have made continued effort seeking strategies to further maximize their financial benefit and minimize risks. For this project, collection agency data were analyzed for the purpose of collecting on past-due accounts receivable balances and seeking strategies to sort through the thousands of records of consumers to increase the re-collectability of the debt. The data were modeled using the methods of Principal Components Analysis, Fisher's Discriminant Analysis, Classification through Logistic Regression, and Binary Decision Tree.

Details

Advisor	Korosteleva, Olga
Author	Yoo, Terrie
Identifier / keyword	Pure sciences; Binary Decision Tree; Credit agency; Debt; Discriminant Analysis; Logistic Regression; Principal Components Analysis
Number of pages	88
ProQuest document ID	1818547490
Publication title	ProQuest Dissertations and Theses
Publication year	2016

Predicting the Success of Running Back Prospects in the National Football League

Merritt, Kevin M.. California State University, Long Beach, ProQuest Dissertations Publishing, 2017. 10287281.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Merritt%2C+Kevin+M.&isbn=9780355097573&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

National Football League team's analysts use statistics in a multitude of ways, including game planning, game day rosters, and incoming talent evaluation. Focusing on the running back position, we attempt to improve upon models designed to predict the future success of incoming collegiate players while introducing some models of our own. Focusing on running backs drafted from 1999 to 2013, we use data from the player's college career, combine workouts, pro day workouts, and physical measurements. Using linear regression, recursive partitioning decision trees, principal component analysis, zero-inflated negative binomial regression, hurdle negative binomial regression, and zero-inflated truncated normal regression, we develop models for three different success criteria: a weighted combination of games played and started, yards per rushing attempt, and career yards from scrimmage.

Details

Advisor	Korosteleva, Olga
Author	Merritt, Kevin M.
Identifier / keyword	Pure sciences; NFL; National football league; Predict; Running back; Statistics
Number of pages	103
ProQuest document ID	1935642292
Publication title	ProQuest Dissertations and Theses
Publication year	2017

Multivariate analysis of proteomic data: Functional group analysis using a global test

Fitzgerald-DeHoog, Lindsay M.. California State University, Long Beach, ProQuest Dissertations Publishing, 2015. 1602759.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Fitzgerald-DeHoog%2C+Lindsay+M.&isbn=9781339186962&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Proteomics is a relatively new discipline being implemented in life science fields. Proteomics allows a whole-systems approach to discerning changes in organismal physiology due to physical perturbations. The advantages of a proteomic approach may be counteracted by the ability to analyze the data in a meaningful way due to inherent problems with statistical assumptions. Furthermore, analyzing significant protein volume differences among treatment groups often requires analysis of numerous proteins even when limiting analyses to a particular protein type or physiological pathway. Improper use of traditional techniques leads to problems with multiple hypotheses testing.

This research will examine two common techniques used to analyze proteomic data and will apply these to a novel proteomic data set. In addition, a Global Test originally developed for gene array data will be employed to discover its utility for proteomic data and the ability to counteract the multiple hypotheses testing problems encountered with traditional analyses.

Details

Advisor	Korosteleva, Olga
Author	Fitzgerald-DeHoog, Lindsay M.
Identifier / keyword	Pure sciences; Biological sciences; Goeman global test; Marine ecology; Multivariate data analysis; Mytilus; Proteomics
Number of pages	135
ProQuest document ID	1734891120
Publication title	ProQuest Dissertations and Theses
Publication year	2015

Structural equation models estimation methods

Nguyen, Hien D.. California State University, Long Beach, ProQuest Dissertations Publishing, 2015. 1602758.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Nguyen%2C+Hien+D.&isbn=9781339186948&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Structural equation modeling (SEM) is a widely used statistical method in the social and behavioral sciences. In its early days, SEM was restricted to linear models among latent variables. This thesis will illustrate the maximum likelihood method for estimating linear models, the product indicator, the two-step methods, and the mixture method for estimating non-linear models. All examples will be executed through the statistical software R. Additionally, examples of bootstrapping will be shown in the context of SEM for the purpose of comparing different estimation methods, performing power analysis, and determining model fit for small and large sample sizes.

Details

Advisor	Korosteleva, Olga
Author	Nguyen, Hien D.
Identifier / keyword	Pure sciences; Equation; Estimation; Modeling; Sem; Structural
Number of pages	108
ProQuest document ID	1734864099
Publication title	ProQuest Dissertations and Theses
Publication year	2015

Machine-Learning Methods for Credit Card Fraud Detection

Woolston, Sarah E.. California State University, Long Beach, ProQuest Dissertations Publishing, 2017. 10602012.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac Learning+Methods+for+Credit+Card+Fraud+Detection&issn=&date=2017-01-01&volume=&issue=&spage=&au=Woolston%2C+Sarah+E.&isbn=9780355219098&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

In order to thwart fraudsters, financial institutions must use current, advanced, customized predictive analytics to protect themselves. Data scientists and statisticians who understand machine learning and statistical methods are in increasingly high-demand and the demand for them is growing each year. Technically, machine learning is a subfield of artificial intelligence whereas statistics is subdivision of mathematics and many believe they only need in depth knowledge of one in order to be a predictive modeler. This fallacy leads to inefficient and/or inaccurate models, and sadly, many industries have not yet realized that the mathematics behind the model is just as important, if not more important, than the computer science needed to implement it. However, some businesses have and this thesis will hopefully help both industry and academia move further along in this direction.

In this thesis, we explore existing methodologies for fraud detection proposed by academic professionals around the globe and illustrate their accuracy, efficiency and reliability on a large dataset downloaded from a public website. The methods analyzed are hidden Markov models (HMM), convolutional neural networks (CNN), and support vector machines (SVM). For each method, we present the history and motivation, theoretical framework, strengths and weaknesses, and numerical examples done in either R or SAS Enterprise Miner.

Details

Advisor	Korosteleva, Olga
Author	Woolston, Sarah E.
Identifier / keyword	Pure sciences; Applied sciences; Classification; Convolutional neural network; Hidden Markov model; Machine learning; Statistics; Support vector machine
Number of pages	100
ProQuest document ID	1954696965
Publication title	ProQuest Dissertations and Theses
Publication year	2017

Cancer: Modeling the distribution of sizes of detectable metastases

Russo, Donna A.. California State University, Long Beach, ProQuest Dissertations Publishing, 2010. 1486390.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Russo%2C+Donna+A.&isbn=9781124247809&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

As life expectancy grows longer in developed nations, the possibility of falling prey to some form of cancer becomes more and more likely. More significantly, while many cancers have become curable or may be sent into remission by palliative treatments, other forms of cancer become metastatic and there is often little that can be done to overcome the disease. Therefore, in an effort to better understand the process of metastatic cancer, researchers in the fields of medicine, biology, epidemiology, clinical investigation, biomathematics, and biostatistics work together to find ways to extend the lives of these cancer patients. Understanding of this process can be facilitated through the use of mathematical, statistical, and stochastic models that approximate progress in the varying stages of this often devastating disease. This thesis will synthesize two existing models for the distribution of sizes of detectable metastases.

Details

Advisor	Korosteleva, Olga
Author	Russo, Donna A.
Identifier / keyword	Health and environmental sciences; Applied sciences; Pure sciences
Number of pages	87
ProQuest document ID	757005752
Publication title	ProQuest Dissertations and Theses
Publication year	2010

User retention and classification in a mobile gaming environment

Ruffin, Michael. California State University, Long Beach, ProQuest Dissertations Publishing, 2014. 1527021.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Ruffin%2C+Michael&isbn=9781321370553&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Game analytics is a fast growing field where game studios are allocating valuable resources to develop sophisticated statistical models to understand user behavior and monetization habits to optimize game play and performance. Game developers' ability to understand user retention allows for game features that will generate high engagement leading to stronger overall monetization and increased lifetimes of players.

One important industry adopted metric is the percentage of users who log back into the game one day after installation, otherwise known as a one-day retention. Although this is an important metric, game studios typically allocate little resources to determining what user transactions are typically conducted on the day of installation that drive a one-day retention.

In this project, we first conduct a cluster analysis in an attempt to uncover meaningful subgroups based on players' transaction history on their first day of installation. Secondly, we use various classification methods including decision trees, logistic regression, and k-Nearest Neighbor algorithm to determine which behaviors are important in identifying whether a new user will return the following day.

Details

Advisor	Korosteleva, Olga
Author	Ruffin, Michael
Identifier / keyword	Pure sciences; Applied sciences; Business intelligence; Cluster analysis; Machine learning techniques; Mobile games; Non-parametric classification; Statistics
Number of pages	52
ProQuest document ID	1630090333
Publication title	ProQuest Dissertations and Theses
Publication year	2014

Data mining techniques for constructing jury selection models

Espy, John. California State University, Long Beach, ProQuest Dissertations Publishing, 2013. 1527548.

Find a copy

Get It CSULB Check for Availability

http://csulb-primo.hosted.exlibrisgroup.com/openurl/01CALS_ULB/01CALS_ULB_SP??url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:Dissertations+%26+Theses+%40+California+State+University%2C+Long+Beac01-01&volume=&issue=&spage=&au=Espy%2C+John&isbn=9781303795879&jtitle=&btitle=&rft_id=info:eric/&rft_id=info:doi/

Abstract

Jury selection can determine a case before it even begins. The goal is to predict whether a juror rules for the plaintiff or the defense in the medical malpractice trials that are conducted, and which variables are significant in predicting this. The data for the analysis were obtained from mock trials that simulated actual trials, with possible arguments from the defense and the plaintiff with ample discussion time. These mock trials were supplemented by surveys that attempted to capture the characteristics and attitudes of the mock juror and the case at hand. The data were modeled using the logistic regression as well as decision trees and neural networks techniques.

Details

Advisor	Korosteleva, Olga
Author	Espy, John
Identifier / keyword	Pure sciences; Decision trees; Logistic regression; Neural networks
Number of pages	48
ProQuest document ID	1513579660
Publication title	ProQuest Dissertations and Theses
Publication year	2013