

BIOSTAT 2017



23rd International Scientific Symposium on Biometrics

ŠIBENIK, CROATIA, 7 - 10 JUNE 2017



BOOK OF ABSTRACTS

EDITED BY: Anamarija Jazbec
Marija Pecina
Zdenko Sonicki
Diana Šimić
Mislav Vedriš

ISSN 1849-434X

CROATIAN
established 1989. **BIO**
METRIC
SOCIETY



HRVATSKO
BIO osnovano 1989.
METRIJSKO
DRUŠTVO

BOOK OF ABSTRACTS

BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

EDITORS:

Anamarija Jazbec
Marija Pecina
Zdenko Sonicki
Diana Šimić
Mislav Vedriš

Zagreb, 2017.

BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

ORGANISER

Croatian Biometric Society



CO-ORGANISER

University of Zagreb



UNDER THE AUSPICES of the
 Croatian Academy of
 Sciences and Arts

INTERNATIONAL SCIENTIFIC PROGRAM COMMITTEE

Anamarija Jazbec (Croatia)
 Marija Pecina (Croatia)
 Franjo Prot (Croatia)
 Zdenko Sonicki (Croatia)
 Diana Šimić (Croatia)
 Bojana Dalbelo Bašić (Croatia)
 Silvije Vuletić (Croatia)
 Lynne Billard (USA)
 Karina Gibert (Spain)
 Annie Morin (France)
 Mislav Vedriš (Croatia)
 Johann Sölkner (Austria)
 Janez Stare (Slovenia)
 John E. Vena (USA)
 Katarina Košmelj (Slovenia)
 Nancy Flournoy (USA)

ORGANIZING COMMITTEE & EDITORIAL BOARD

Anamarija Jazbec
 University of Zagreb, Faculty of Forestry
 Marija Pecina
 University of Zagreb, Faculty of Agriculture
 Zdenko Sonicki
 University of Zagreb, School of Medicine,
 School of Public Health „Andrija Štampar“
 Diana Šimić
 University of Zagreb, Faculty of Organization
 and Informatics
 Mislav Vedriš
 University of Zagreb, Faculty of Forestry

21st SCHOOL OF BIOMETRICS

Snehalata Huzurbazar
 University of Wyoming, USA

KEYNOTE LECTURE

Janez Stare
 University of Ljubljana, Slovenia

PUBLISHER:

Croatian Biometric Society
 c/o University of Zagreb, School of Medicine,
 School of Public Health “Andrija Štampar”
 Rockefellerova 4, HR-10000 Zagreb, Croatia
www.hbmd.hr

FOR THE PUBLISHER:

Professor Anamarija Jazbec, Ph.D., President
 University of Zagreb, Faculty of Forestry
 Svetošimunska 25, HR-10000 Zagreb, Croatia
hbmd@hbmd.hr

ISSN 1849-434X

FULL REFERENCE:

Jazbec A, Pecina M, Sonicki Z, Šimić D, Vedriš M, eds. (2017) Book of Abstracts BIOSTAT 2017 – 23rd International Scientific Symposium on Biometrics, Šibenik, 7 - 10 June 2017. Zagreb: Croatian Biometric Society. ISSN 1849-434X

VENUE:

Solaris Hotel Resort
 Hoteli Solaris 86
 HR - 22000 Šibenik, Croatia
www.solarishotelsresort.com



BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

PREFACE	1
PROGRAM	2
ABSTRACTS	
21 st SCHOOL OF BIOMETRICS	5
Snehalata Huzurbazar University of Wyoming, USA	
KEYNOTE LECTURE	7
Janez Stare University of Ljubljana, Slovenia	
CONTRIBUTED SESSIONS	9
AUTHOR INDEX	29



BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

Welcome to the Book of Abstracts of BIOSTAT 2017 – The 23rd International Scientific Symposium on Biometrics organized by the Croatian Biometric Society and co-organized by the University of Zagreb – held in Šibenik, Croatia, June 7 – June 10, 2017.

It is our great honor to announce that this year's BIOSTAT Symposium is being held under the auspices of the Croatian Academy of Sciences and Arts.

BIOSTAT 2017 comprises 21st School of Biometrics, keynote lecture and contributed papers. This brochure contains 22 abstracts written by 46 authors from six countries – Bosnia and Herzegovina, Croatia, Netherlands, Slovenia, United Kingdom and United States of America.

BIOSTAT symposium is an intrinsically interdisciplinary meeting accepting papers from a wide array of research disciplines at the intersection of data sciences and life sciences as evidenced by this BIOSTAT 2017. After organizing 18 Schools and 20 BIOSTATs within ITI – International Conference on Information Technology Interfaces – this is the third time that BIOSTAT and the School are held as independent events.

Symposium is taking place in Solaris Hotel Resort near Šibenik – the oldest native Croatian city on the Adriatic located in the most protected natural harbor in picturesque and indented bay at the mouth of the Krka River. Created initially as an ancient Croatian *castrum*, a fortification or encampment, at the bottom of St. Michael's Fortress that still dominates the town, Šibenik was mentioned for the first time in 1066, in a document issued by the most important Croatian ruler - the king Petar Krešimir IV. Šibenik obtained the status of a native Croatian town in 1290, when the Diocese of Šibenik was established.

We expect that this year again our friendship, exchange of ideas and interchange of knowledge will contribute to the progress of Biometric science.

Invited lecturer of the 21st School of Biometrics is

Professor **Snehalata Huzurbazar**, PhD (University of Wyoming, USA) with the topic:
Human Microbiome Data Analysis: Overview and Statistical Problems

This year keynote lecture is

Professor **Janez Stare**, PhD (University of Ljubljana, Slovenia) with the topic:
R² – a simple statistic simply misunderstood. Still.

We hope to share with you pleasant moments, lively discussions, successful networking and forge new partnerships and friendships.

Editors:
Anamarija Jazbec, Marija Pecina, Zdenko Sonicki, Diana Šimić and Mislav Vedriš
Zagreb, June, 2017

BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

PRELIMINARY PROGRAM BIOSTAT 2017

	Wednesday, June 7	Thursday, June 8	Friday, June 9	Saturday, June 10
9:00-9:30		Opening Ceremony		
9:30-10:30		School of Biometrics Part 1.	<i>Contributed session</i>	<i>Contributed session</i>
10:30-11:00		COFFEE BREAK	COFFEE BREAK	COFFEE BREAK
11:00-12:00		School of Biometrics Part 2. Snehalata Huzurbazar	Keynote Lecture Janez Stare	<i>Contributed session</i>
12:00-13:00			<i>Contributed session</i>	Closing Ceremony
13:00-14:00		LUNCH BREAK	LUNCH BREAK	LUNCH BREAK
14:00-15:30		<i>Contributed session</i>		
15:30-16:00		COFFEE BREAK		
16:00-17:00		<i>Contributed session</i>		
17:00-18:00		ISPC Meeting	Excursion and Dinner	
18:00-19:00				
19:00-20:00	Registration			
20:00 -	Welcome Cocktail			

BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
 ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

SYMPOSIUM PROGRAM In Detail

Wednesday June, 7	19:00	Registration & Welcome Cocktail		
	20:00			
Thursday, June, 8	9:00 – 10:30	<i>Chair: Zdenko Sonicki</i> Snehalata Huzurbazar	School of Biometrics: Human Microbiome Data Analysis: Overview and Statistical Problems Part 1	
	10:30 – 11:00	COFFEE BREAK		
	11:00 – 12:45	<i>Chair: Zdenko Sonicki</i> Snehalata Huzurbazar	School of Biometrics: Human Microbiome Data Analysis: Overview and Statistical Problems Part 2	
	13:00 – 14:00	LUNCH BREAK		
	14:00 - 15:30	<i>Chair: Diana Šimić</i>	Contributed session	
		AUTHOR	TITLE	
		Lynne Billard	Waiting Time Approach for Compartment Models: Impact of HIV-AIDS on Insurance/Health Costs	
		Alexander V. Alekseyenko	Multivariate methods for analysis of microbiome data	
		Jihad S. Obeid, Alexander V. Alekseyenko, Jie Zheng, Anna Maria Masci, Yongqun "Oliver" He, DVM,	An Ontology for Host-Microbiome Interactions	
	Cameron Miller, Benjamin Neely, Richard Drake, Elizabeth Hill	A Two-Part Process Convolution Model to Analyze Semicontinuous, Spatially-Referenced MALDI Imaging Data		
15:30 – 16:00	COFFEE BREAK			
Friday, June, 9	9:00 - 10:30	<i>Chair: John Vena</i>	Contributed session	
		AUTHOR	TITLE	
		Paul A. Thompson, Norm Matloff	Combating image fraud by developing better tools for image manipulation	
		Elizabeth G. Hill, Kent Armeson and Bonnie Martin-Harris	A Bayesian Ordinal Latent Class Model for Dysphagia Severity	
		Milan Rimac, Jelena Roganovic, Zdenko Sonicki	Spatio-temporal model of the most common childhood cancers occurrences	
	Borna Pleše, Damir Ivanković	Analysis of data quality dimensions in Croatian public health registries and suggestions for improvement		
10:30 – 11:00	COFFEE BREAK			

	11:00 – 12:00	<i>Chair: Anamarija Jazbec</i> Janez Stare	Keynote Lecture: R2 – a simple statistic simply misunderstood. Still.
	12:00 – 13:00	<i>Chair: Janez Stare</i>	Contributed session
		AUTHOR	TITLE
		Lavoslav Čaklović	Statistical Analysis of Cognitive Domain Taxonomy Table
		Franjo Prot, Sara Prot	Multivariate Comparisons of Alternate Forms of Composite Measuring Instrument Designed by Clustering Approach
		Falak Hrvoje, Perišić Ana, Urbanc Jadranka	Is the answer blowing in the wind? The association between mental-behavioral disorder admissions and wind parameters
	13:00 – 14:00	LUNCH BREAK	
	14:00 – 15:00	ISPC Meeting	
18:00 – 24:00	Excursion and Dinner		
Saturday, June 10	9:00 - 10:30	<i>Chair: Marija Pecina</i>	Contributed session
		AUTHOR	TITLE
		Želimir Kurtanjek	Statistical analysis of yeast whole genome expression under variable nutrient limitations
		Jasminka Dobša, Henk A. L. Kiers	Visualization of data in lower dimensional space
		Gusić Jelena, Šimić Diana, Mina Đorđević	Psychometric properties of a Croatian version of student experience survey
		Marijana Žunić, Krunoslav Teslak	Understanding small-scale private forest owners management behavior-application of confirmatory factor analysis
	10:30 – 11:00	COFFEE BREAK	
	11:00 – 12:30	<i>Chair: Katarina Košmelj</i>	Contributed session
		AUTHOR	TITLE
		Bojana Dalbello Bašić	Is there more common ground between statistics and machine learning than what meets the eye?
		Borut Bosančić, Marija Pecina, Nikola Mičić	Main Effect Meta Principal Component Analysis (ME-MetaPCA) as the Tool of Choice for Processing Typical Horticultural Metadata
		Martina Tijardović, Tomislav Dubravac, Sanja Perić	Analysis of influence of seedling type and quality on the out planting success after disturbances in Norway spruce forest culture
		Mislav Vedriš, Anamarija Jazbec	Why (not) study longer? Analysis of duration of undergraduate studies at Faculty of Forestry UNIZG
	12:30 - 13:00	Closing Ceremony	
13:00 – 14:00	LUNCH BREAK		

follow us at www.hbmd.hr

ABSTRACT



BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

21st SCHOOL OF BIOMETRICS

BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

21st SCHOOL OF BIOMETRICS



Snehalata Huzurbazar

University of Wyoming,
Department of Statistics,
Laramie, WY 82071, USA

lata@uwyo.edu

http://www.uwyo.edu/statistics/facultyandstaff/huzurbazar_snehalata.html

Topic:

Human Microbiome Data Analysis: Overview and Statistical Problems

Abstract:

By latest estimates, the human body contains about half human cells and half bacterial cells, though previous estimates have this ratio as skewed as 1:10. The bacteria form ecosystems in the human body and are referred to as the human microbiome, which has become increasingly important for various aspects of human health. The main source for microbiome data is from sequencing the 16S rRNA gene. The resulting data are highly sparse, high-dimensional counts usually with fewer samples than number of (bacterial) taxa, namely, $n < p$. My experience comes from working with the vaginal microbiome along with covariate data from female subjects as part of a U.S. National Institutes of Health Project. As we worked with such data, we used the opportunity to examine common first steps such as filtering and normalization, and developed some proposals for improvement. We then explored meaningful visualizations that can lend biological insights for such data. In the process, we compared various ordination methods that are commonly used by biologists for analyzing these data. Currently, we are examining methods for longitudinal analysis for such data collected over a handful of repeated visits.

Using my experience, the lectures will be structured to give the audience an overview of the methods currently used to analyze human microbiome data, and also to point towards the open problems where the statistical sciences could make contributions.

ABSTRACT



BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, **7 - 10 JUNE 2017**

KEYNOTE LECTURE

BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

KEYNOTE LECTURE



Janez Stare

University of Ljubljana, Faculty of Medicine,
Institute for Biostatistics and Medical Informatics,
Ljubljana, Slovenia

janez.stare@mf.uni-lj.si

<http://ibmi.mf.uni-lj.si/janez/index.php?lang=en-GB>

Topic:

R^2 – a simple statistic simply misunderstood. Still.

Abstract:

The recent misuse of R^2 in a paper in Science is a good reminder of the fact that this, apparently simple, statistic is still often misunderstood.

The Science paper is an example of reading too much into R^2 , and the same goes for its interpretation as a measure of goodness of fit. The other side of misunderstanding of R^2 is represented by its many criticisms, most of them unfounded, and those left could be addressed to almost any other statistic. Adding to these the fact that all software packages report a meaningless R^2 in the case of linear regression with zero constant, it becomes obvious that there is a lot of confusion out there when it comes to understanding the message communicated by R^2 . The most persistent criticism is about the fact that R^2 depends on distribution of covariates, and the most persisting misunderstanding is about R^2 measuring goodness of fit. In this talk I will briefly revisit the criticisms directed towards R^2 and explain the falseness of the accusations, and then focus on the two points, mentioned above. My message is: a) dependence on the distribution of covariates is not a property, specific to R^2 , and b) R^2 is not a measure of goodness of fit. And, to further emphasize this last point, I will conclude with a brief explanation of how to formally test goodness of fit in linear regression.

ABSTRACTS



BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

CONTRIBUTED SESSIONS

Title:	Waiting Time Approach for Compartment Models: Impact of HIV-AIDS on Insurance/Health Costs
Authors:	Lynne Billard
Affiliation:	University of Georgia
Abstract:	<p>Traditionally, epidemic processes have focused on establishing systems of differential-difference equations governing the number of individuals at each stage of the epidemic. Except for simple situations such as when transition rates are linear, these equations are notoriously intractable mathematically. We adopt a waiting time approach for multi-stage compartmental models. The possibility that individuals can die at any stage from non-disease related causes is also included. The model is used to study the impact of HIV-AIDS on insurance and health-care costs.</p>

Title:	Multivariate methods for analysis of microbiome data
Authors:	Alexander V. Alekseyenko
Affiliation:	Medical University of South Carolina
Abstract:	<p>Microbiome data comes as multivariate abundance profiles of taxa across samples, often referred to as communities. Drawing from numerical ecology literature, microbiome analysis relies on distance based techniques, know as beta diversity analysis. The data are summarized as pairwise, sample-to-sample, representation using distance metrics that capture various useful aspects of the data, such as phylogenetic relationships, compositions, correlations. Distance-based approaches allow for effective visualization of these multivariate data using lower-dimensional projection techniques, such as multidimensional scaling or principal coordinates analysis. These descriptive methods are often augmented by more formal hypothesis-testing analytics. Suppose observations are grouped into k discrete categories. Inference about association of the grouping with the microbiota can be obtained using non-Euclidean permutational multivariate analysis of variance (PERMANOVA). This approach tests the omnibus hypothesis of difference in microbial composition/abundance between the categories. We have previously shown that PERMANOVA is not robust to heteroscedasticity and sample imbalance in the data, features that are unfortunately common in microbiome datasets. To overcome this challenge in two-sample situations, we have derived an alternative approach by distance-based multivariate extension of the Welch t-test, T_{w2}. We have shown that the resulting permutational test has the desired type I error and power properties. More recently, we have extended this solution to an arbitrary number of categories. We do so by considering heteroscedastic Welch ANOVA statistic W^* and deriving distance-based calculation, W_{d^*} to extend it to the multivariate case. We show that T_{w2} is a special case of this more general solution. We further empirically evaluate the performance of W_{d^*}. In this talk, we present these results as well as describe the general approach to multivariate analysis of microbiome datasets.</p> <p>This work will be presented by Prof. John Vena.</p>

Title:	An Ontology for Host-Microbiome Interactions
Authors:	Jihad S. Obeid, MD ^{1*} , Alexander V. Alekseyenko, PhD ¹ , Jie Zheng, PhD ² , Anna Maria Masci, PhD ³ , Yongqun "Oliver" He, DVM, PhD ⁴
Affiliation:	¹ Medical University of South Carolina, ² University of Pennsylvania, ³ Duke University, ⁴ University of Michigan *Presenter
Abstract:	<p>Microbiome analyses can be diverse, complex and data intensive. Reproducibility of experiments depend on the use of consistent procedures across different settings. Data across different institutions can only be shared when using identical or similar protocols and data models. Although there are existing standards for representing microbial organisms, the terminology of the interactions between microbiomes and hosts along with associated biological processes has not been harmonized. We present a work in progress on the development of a community-based Ontology of Host-Microbiome Interactions (OHMI) through a multi-institutional collaboration. Given the large amounts of microbiome data being generated, the goal of OHMI is to create a common ontology to represent a variety of host (e.g., human and mouse), microbiome, host-microbiome interactions under different conditions, protocols, resulting data and possible analyses. For example, the hierarchies of bacterial and fungal microbiota are represented under the super class of microbiome. Doing so will permit improved sharing and pooling of data, analyses and results. Development will follow the Open Biomedical Ontologies Foundry (http://www.obofoundry.org/) principles. Several use cases have been identified, including those related to prokaryotic 16S rRNA sequence analysis, antibiotics effects on microbiome, and gut microbiome influences on human diseases (e.g., obesity, diabetes, and cancer).</p>

Title:	A Two-Part Process Convolution Model to Analyze Semicontinuous, Spatially-Referenced MALDI Imaging Data
Authors:	Cameron Miller ¹ , Benjamin Neely ² , Richard Drake ³ , Elizabeth Hill ¹
Affiliation:	¹ Department of Public Health Sciences, Medical University of South Carolina ² Chemical Sciences Division, National Institute of Standards and Technology ³ Department of Cell and Molecular Pharmacology, Medical University of South Carolina
Abstract:	<p>Matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance imaging mass spectrometry (IMS) technology allows researchers to measure the abundance of ionized fragments over a two-dimensional space. Despite advances in IMS technology, methods used to analyze such data have lagged. In particular, the variability in IMS data can be attributed to both spatial and random sources. Additionally, the frequency of masses with high proportions of zero peak intensity values is often quite large. To address these issues, we propose a log-linear regression model facilitating group-level comparisons of ionized fragment abundance, which further accounts for both the data's spatial structure and semicontinuous nature via a two-part model. We evaluate our model using simulated data and compare performance to a naïve analysis using a two-sample t-test. We then use our approach to identify glycans significantly associated with stage 1 renal cell carcinoma (RCC) using IMS data obtained from tissue microarray RCC samples.</p>

Title:	Combating image fraud by developing better tools for image manipulation
Authors:	Paul A. Thompson, Ph.D., PSTAT(R) ^{1,2} Norm Matloff, Ph.D. ³
Affiliation:	¹ Sanford Research ² University of South Dakota ³ University of California, Davis
Abstract:	<p>Research fraud is a continuing problem in the scientific enterprise. Image fraud continues to be a large part of the fraud problem. Recent studies have found increases in image fraud. The approach for working with image fraud is a post-hoc forensic image examination approach. While effective, this does not address the root cause, improper preparation of images. Reproducible research methods increasingly indicate that structured, scripted preparation of documents is key to reproducible outcomes. To repeat the result, the process must be clearly defined. In image preparation, the approach is an interactive one using a graphical user interface to work with images. A “journaling process” (preparing an executable concurrent log of manipulations) is under development for open-source image manipulation tools. Journaling will produce a log that can be examined, and used to demonstrate that the manipulations performed prepared the submitted image. The journaling process was proposed for GIMP and R. GIMP will be modified to output ImageMagick code for all operations that are performed in the GUI. In addition, it will be modified to output script-fu commands. GIMP will also be modified to perform only appropriate scientific image manipulations. The R tool, called ShinyImage, uses EImage code to both display images under manipulation, and output manipulation code. The methodological development will enhance legitimate scientific image manipulation. It will make image manipulation transparent, and will clearly exhibit the changes that have been used. Image fraud methods which assist users in appropriate preparation are better than those which find inappropriate preparation.</p>

Title:	A Bayesian Ordinal Latent Class Model for Dysphagia Severity
Authors:	Elizabeth G. Hill ¹ , Kent Armeson ¹ and Bonnie Martin-Harris ²
Affiliation:	¹ Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC ² Department of Communication Sciences and Disorders, Northwestern University, Chicago, IL
Abstract:	<p>The modified barium swallow study (MBSS) – an exam in which a patient swallows solids and liquids of varying textures, viscosities and volumes (collectively known as 'tasks') – is used to identify bolus consistencies that a patient with dysphagia (swallowing disorder) can safely ingest. The Modified Barium Swallow Impairment Profile (MBSImP) is an instrument used to identify specific physiologic components of swallow impairment using the MBSS video-fluoroscopic image. MBSImP components are ordinally scored for each task across oral, pharyngeal and esophageal components. For a given component, the maximum (worst) score across all tasks is used to guide clinical intervention. We propose a Bayesian ordinal latent class model to discover latent classes of dysphagia severity based on both task-level MBSImP component scores and maximum scores, where the conditional likelihood for the latter is formulated using properties of the distribution of the maximum order statistic. We demonstrate the utility of our approach to identify dysphagia severity classes and explore associations between latent class and known clinical measures of dysphagia.</p>

Title:	Spatio-temporal model of the most common childhood cancers occurrences
Authors:	Milan Rimac, MD, MSc ^{1*} ; Jelena Roganovic, MD, PhD ¹ ; Zdenko Sonicki, MD, PhD ²
Affiliation:	¹ Clinical Hospital Center Rijeka, University of Rijeka School of Medicine, ² „Andrija Stampar“ School of Public Health, University of Zagreb School of Medicine; *Presenter
Abstract:	<p>Childhood cancer is age-defined subgroup of malignant diseases distinguished by low incidence, specific tumorigenesis and unknown etiology. Epidemiological studies of spatial and temporal distributions as well as spatio-temporal analysis of childhood cancer occurrences aimed at portraying population genetic landscape and depicting affecting environmental factors could be important in clarifying childhood cancer etiology. Previous studies have documented distinctiveness of researched populations and suggested spatio-temporal clustering of diseases as well as occurrence periodicity of particular malignancies.</p> <p>We present ongoing multicenter study aimed to investigate spatial, temporal, and spatio-temporal distribution of the most frequent childhood cancers occurrences in the Republic of Croatia. Clinical data for children aged 0-14 years diagnosed with first malignancy in a period 2001-2015 at each of the five pediatric oncology departments in Croatia would be collected. Respective addresses at diagnosis would be geocoded with the tool provided by GDİ GISDATA (Zagreb, Croatia).</p> <p>Spatial data from Central register of spatial units (State geodetic administration), and 2001-2011 population censuses data (Croatian bureau of statistics) would be used. Global and regional age-specific incidence rates as well as age-standardized incidences would be assessed, and their spatial and temporal distribution analyzed. Geographic information system tool ArcGIS 10.5 (ESRI, Redlands, CA, USA) would be used for data presentation and further analysis.</p> <p>Results of this study would be a contribution to the clarification of regional appearances of childhood cancer and possible implications of environmental and population risk factors.</p>

Title:	Analysis of data quality dimensions in Croatian public health registries and suggestions for improvement
Authors:	Borna Pleše, Damir Ivanković
Affiliation:	Croatian Institute of Public Health (Hrvatski zavod za javno zdravstvo)
Abstract:	<p>The need for efficient data governance and use has become an inevitable necessity in modern healthcare systems which continue to expand in complexity, resources and variety of services provided and needs identified. However, there is a significant disconnect between the optimal process of data collection, sharing and use to monitor and improve these changes, and the practical priorities of the stakeholders involved at all levels of the system.</p> <p>The frequent emphasis on secondary use of data is both logical and hindering for the continual improvement of available data, since there are no accepted/validated criteria to assess governance, quality and safety/security when deciding on secondary use (providing that the data itself is at once accurate, representative and interpretable for the expressed purposes of analysis). The current channels of data collection and use often fall on inefficient systems of collection (paper-based reporting), use (skewed reporting based on data which has not been fully validated or is incomplete), and infrequent revision of data collection and access procedures. However, improvements of data quality can improve both secondary use and greatly optimize and expand the utility of primary data (e.g. surveillance, outcomes monitoring, HTA).</p> <p>The increasing production of data can be counterproductive when poorly managed, so rational scaling and continuous assessment of data quality within the system is needed to ensure rational use of both technical and human resources. This is a continuous process and involves specific as well as universal best practice approaches, from adapting governance models to education of actors within the system both on the data provider and user sides.</p> <p>A case study of the Croatian national public health information system (NAJS) is a good example of how shifting perspective from a disconnected to an integrative system presumes several levels of interoperability to be established between all stakeholders and challenges met.</p> <p>Using this example, several levels of data quality in public healthcare registries can be investigated as well as the process of data collection and how the two can be optimized. The data itself can be assessed per data quality dimensions of timeliness, relevance, accuracy and completeness.</p> <p>Practical recommendations are proposed as methods of mitigating risks of unsatisfactory data quality in the short term, such as moving from population based reporting to validated sample-based data analysis as well as long term recommendations for further study and practical steps to systematically improve healthcare data for public health purposes.</p>

Title:	Statistical Analysis of Cognitive Domain Taxonomy Table
Authors:	Lavoslav Čaklović
Affiliation:	Faculty of Natural Sciences – Department of Mathematics
Abstract:	<p>Taxonomy table of the cognitive domain has 4 categories for knowledge and 6 categories for processes. The complexity value of the cell in that table depends upon the 'distance' from the left upper corner. Our idea is to calculate this complexity value using the contingency table obtained in the following way, in fact there are several ways to do that. We asked each student of the course in Game Theory to evaluate his own knowledge (a day before the exam) on the taxonomy table (TT) by putting a point in one or several cells of the table.</p> <p>Correspondence Analysis (CA) of that contingency table gave a strong nonlinear association which was again analysed by the Detrended CA (DCA). The final result is that the association is one dimensional which gives rise to some interesting questions about the validity of the proposed taxonomy.</p> <p>Furthermore, the scores obtained by DCA may be used to generate the metric of the TT which evaluates each cell of the table. This metric may be used to evaluate the students abilities and outcomes. Further experiments and analysis have to be done.</p>

Title:	Multivariate Comparisons of Alternate Forms of Composite Measuring Instrument Designed by Clustering Approach
Authors:	Franjo Prot ¹ , Sara Prot ²
Affiliation:	¹ Faculty of Kinesiology, University of Zagreb, Croatia ² School of Psychological, Social and Behavioural Sciences, Coventry University United Kingdom, Contact: ac2215@coventry.ac.uk
Abstract:	<p>Key words: clustering, alternate forms, measurement properties</p> <p>Representative collection of 78 items of attitudes toward science and scientists had been produced as a basic items pool for construction consistent scale(s). Items were presented to 322 subjects for evaluation. Items were appropriately oriented to the logically positive direction. Items correlation exhibited only 5% negative correlation due to expected complexity of attitude relations of some items. Number of items for the further analysis is reduced to 60 items corresponding to items which had value greater than 0.45 on the first principal component of initial set of items. In that way 18 items were eliminated from the pool. Standard psychometric analysis for all 60 items exhibited reliability Alpha = 0.97 as an upper bound for shorter collection of items. Correlations between selected items were all positive what enables the next stage of analysis with the aim to constrict alternate forms. Centroid cluster analysis on correlations of 60 items produced 29 pairs of similar items, i.e. cluster on the first level. Only two items were found isolated on the first basic level. One of the item from the pair was allocated to the first form and the remaining one to the second form. Two times two forms were produce systematically select the items from the pairs in 1,2,1,2, .. manner for the first pair of alternate forms and 1,2,2,1,... manner for the second pair of alternate forms. The first pair of 29 items of alternate forms showed reliabilities Alpha 0.95 respectively, and the second pair of 29 items of alternate forms showed reliabilities Alpha 0,95 respectively. Reducing first and second pairs of alternate forms to 15 items reduced reliabilities to 0.95 and 0.93 respectively for the first pair, and 0.94 and 0.94 respectively for the second pair of alternate forms. Multivariate relations of the first pair of two 15 items alternate forms are showing the first canonical correlation of 0,970, the first quasi canonical correlation of 0.962 and the first nuclear correlation of 0.969. Multivariate relations of the second pair of two 15 items alternate forms are showing the first canonical correlation of 0,965, the first quasi canonical correlation of 0.958 and the first nuclear correlation of 0.963</p>

Title:	Is the answer blowing in the wind? The association between mental-behavioral disorder admissions and wind parameters
Authors:	Falak Hrvoje ¹ Perišić Ana ² Urbanc Jadranka ³
Affiliation:	¹ Clinical Hospital Dubrava ² Polytechnic of Šibenik ³ Department of Emergency Medicine Krapina Zagorje County
Abstract:	<p>Recognition that human health and disease can be affected by weather conditions probably predates written history. The effect of wind was understood as particularly strong, and for this reason some countries had prohibited any important decision-making, or even pardoned crimes committed, during certain wind types. This work examines the relationship between daily wind parameters and the number of medical admissions of patients who are diagnosed with mental and behavioral disorders or with external causes of morbidity and mortality. Data for this research was collected from April 2012 to March 2013 and encompasses 1299 patients admitted through Department of Emergency Medicine Krapina-Zagorje County, Croatia. In particular, this study covers 7 groups of diagnoses, classified according to International Classification of Diseases (ICD), and wind parameters such as direction, strength and speed. Moreover, the seasonality of each diagnosis group and group differences in demographic characteristics of patients are examined. Correspondence analysis and logistic regression were applied revealing both, significant association between admissions of some diagnosis groups and wind parameters, and significant seasonal variation of diagnoses, where similar seasonal patterns across some diagnosis groups were encountered.</p>

Title:	Statistical analysis of yeast whole genome expression under variable nutrient limitations
Authors:	Želimir Kurtanjek
Affiliation:	University of Zagreb, Faculty of Food Technology and Biotechnology*, Zagreb, Croatia zelimir.kurtanjek@gmail.com
Abstract:	<p>Mathematical analysis and statistical inference in whole genome studies is a difficult and very challenging task. The problem usually is due to several order of magnitude larger dimension in genome data (usually expression levels) than observed biological features. Here are presented results of yeast <i>Saccharomyces cerevisiae</i> whole genome expression profiles under various nutritional limitations under steady state stress conditions (limitations in nutrients (glucose, ammonium, sulfate, phosphate, uracil, and leucine) in a continuous flow through culture (chemostat). Studied is influence of the cell specific growth rate in the range from 0.005 to 0.3 h⁻¹. on the expression levels¹. The experimental data are available in open data repository². Applied are linear and nonlinear models for statistical inference of the nutrients and growth rates on the clusters of genes. The gene clusters are inferred by the linear models with sparse principal components, sparse Fisher discrimination, and the elastic nets. For inference by nonlinear model used is ensemble of decision trees with the extreme gradient boosting³. Observed is that the largest gene cluster has expression linearly proportional to cell growth rate, but also are identified the specific clusters with striking differences in expressions due to limitations by ammonium, sulfate and phosphate. The main objective in this work is comparative analysis of the linear and nonlinear inference algorithms for identification of expression of gene cluster under the experimentally designed biological stresses.</p> <p>[1] M. Brauer, et.al., <i>Mol. Biol. Cell.</i> 19(1) 2008, 352-367 [2] http://varianceexplained.org/files/Brauer2008_DataSet1.tds [3] B. Efron, T. Hastie, "Computer Age Statistical Inference", Cambridge University Press, 2016 [4] R Development Core Team, R: A language and environment for statistical computing, R, Vienna, Austria. www.R-project.org 2016 [5] T. Chen, et al., Extreme Gradient Boosting. R package version 0.4-3. https://CRAN.R-project.org/package=xgboost</p> <p>*retired</p>

Title:	Visualization of data in lower dimensional space
Authors:	Jasminka Dobša, Henk A. L. Kiers
Affiliation:	Faculty of Organization and Informatics, University of Zagreb, Croatia Department of Psychology, University of Groningen, Netherlands
Abstract:	<p>Reduction of dimensions of data representation can be beneficial for diverse statistics or machine learning tasks. It can be achieved by selection of existing variables or by extraction of new variables. Our aim is to use methods for dimensionality reduction by extraction of new variables in order to represent data in a way which will enable better performance of clustering and automatic classification of objects. We will use methods of Reduced k-means, Factorial k-means and propose a modified method of Reduced k-means with penalization to project data in lower dimensional space and visualize it. Reduced k-means and Factorial k-means methods are methods that aim at reduction of variables by capturing the clustering structure present in the data. The aim of research is to test usefulness of these methods for clustering and classification of objects by visualization in lower dimensional space.</p>

Title:	Psychometric properties of a Croatian version of student experience survey
Authors:	Gusić Jelena ¹ , Šimić Diana ¹ , Mina Đorđević ² Jelena.Gusic@foi.hr; Diana.Simic@foi.hr; Mina.Dordevic@azvo.hr
Affiliation:	¹ University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia ² Agency for Science and Higher Education (ASHE), Zagreb, Croatia
Abstract:	<p>Background. Student experience and engagement are becoming increasingly important aspects of quality assurance in higher education. National level student surveys have been established in several countries (NSS in UK, NSSE in USA, ISSE in Ireland, CEQ in Australia and China etc.). Agency for Science and Higher Education (ASHE) has developed a similar questionnaire on student experience and engagement in higher education. The purpose of this study is to test reliability of this questionnaire and explore its factor structure.</p> <p>Methods. In 2016 students who had enrolled in an undergraduate study program in academic year 2012/2013 were contacted by e-mail and asked to fill in an online questionnaire. The questionnaire comprised 288 questions on various aspects of student experience and engagement, and some demographic data. We selected 98 items related to student attitudes organized in 19 sections (dimensions) and measured on a 7 point ordinal scale ranging from strongly disagree to strongly agree. Out of 2432 students who accessed the online survey, 695 completed all analyzed questions. We used exploratory factor analysis (EFA) to identify number of dimensions and their structure.</p> <p>Results. Descriptive statistics ranged from 2.97 to 5.71 (mean), 1.38 to 2.20 (standard deviation), 2 to 6 (median), -1.36 to 0.66 (skewness), -1.36 to 1.30 (kurtosis). All items covered full range of values from 1 to 7. Overall KMO MSA was 0.97 with item MSA ranging between 0.70 and 0.98. Principal axis factor analysis on Pearson correlation matrix with oblimin rotation identified 14 dimensions (parallel analysis and BIC) explaining 58% of total variance. Cronbach alpha of the whole scale was 0.98 with ITCC ranging from 0.11 to 0.78. Only one item had ITCC<0.3, and 10 items had ITCC<0.4. There was a good correspondence between factors and questionnaire sections. Items from one or two sections loaded on each factor, and some sections were divided between two factors.</p> <p>Conclusion. EFA helped identify items that should be rephrased or removed from the questionnaire. We will use these results to design a shorter version of student experience survey. Further research on association between this scale and performance in studying, as well as collection of data on a representative sample of student population are needed before we can use survey results in policy making.</p> <p>Keywords: student experience survey, exploratory factor analysis</p>

Title:	Understanding small-scale private forest owners management behavior-application of confirmatory factor analysis
Authors:	Zunic Marijana, MEng in Forestry Teslak Krunoslav, Assistant Professor
Affiliation:	University of Zagreb, Faculty of Forestry Department of Forest Inventory and Management
Abstract:	<p>Privately-owned small-scale forest lands are important source of timber and ecosystem services both worldwide and in Croatia. Due to complex and diverse private forest owners objectives it is difficult to estimate potential wood supply from private forest lands. The present study was designed based on the theory that private forest owner decisions-making consists of internal cognitive processes which determines behavior in social situations. To understand how private forest owners behave we conducted a 1007 telephone interviews with randomly selected private forest owners where they were asked to asses 19 statements defining forest management (expressing their own opinion) using a five-point Likert scale. To mesure this variables we applied a confirmatory factor analysis (CFA) and tested whether the data fit a hypothesized measurement model based on theory and previous research. The LISREL 8.80 computer program was employed to conduct a confirmatory factor analysis and assess the three factor model. Several goodness-of-fit were used to determine how well the model fits to the data: Chi-squared test, the Root mean square error of approximation (RMSEA), the Comparative fit index (CFI), and the Standardised root mean square residual (SRMR). Results indicated a good fit between the model and the observed data. The model also serves as an important basis for further research of private forest owner management behavior.</p>

Title:	Is there more common ground between statistics and machine learning than what meets the eye?
Authors:	Bojana Dalbelo Bašić
Affiliation:	University of Zagreb Faculty of Electrical Engineering and Computing Text Analysis and Knowledge Engineering Lab Unska 3, HR-10000 Zagreb, CROATIA
Abstract:	<p>Machine learning is a subfield of artificial intelligence and is concerned with learning from experience (e.g. data), which means building computational models without being explicitly programmed. Statistics is a subfield of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data. On the general level, machine learning and statistics are both concerned with collecting and analyzing data, but these two fields face similar problems, in similar but not identical ways.</p> <p>Much older and more traditional field of statistics was developed in the statistical (mathematical) department, while machine learning is developed in the computer science department. So these two communities have different backgrounds, they go to different conferences, publish in different journals and they are putting different things as the most important in data analysis.</p> <p>One example of such different background is terminology. There are examples where two communities use different terms for the same notion (e.g. training sample are data, weights are parameters, learning is fitting) and even vice versa, where the same term is used for different notions (e.g. biometrics - different fields in computer science and statistic).</p> <p>Some of very established and successful methods in data analysis like support vector machines and neural networks do not come from the field of statistics, as they are not probabilistic models in their nature. Also, some of the big statistical results have come from outside the field of statistics, like boosting.</p> <p>Although many statisticians were critical to the machine learning community, accusing them of misuse the methods and lack of a fully developed concept of inference - beyond predictive accuracy (and even vice versa), statisticians like Brieman, Tibshirani, Friedman, Hastie, Efron and others established their leadership in both fields, working on large scale inference problems, contributing so that there is now much less difference and misunderstanding between the computer scientists and statisticians then before.</p> <p>Many new field/subfields exist and new have appeared lately, bringing more entropy and fuzziness to definitions of borders of all of them: data science, big data, KDD, data mining, pattern recognition, etc.</p> <p>The question I will try to answer is: Do we have to define these borders more clearly or can we learn from each other, can we bring different views together to give us a synergetic effect, maybe under a new integrating field named <i>data science</i>?</p>

Title:	Main Effect Meta Principal Component Analysis (ME-MetaPCA) as the Tool of Choice for Processing Typical Horticultural Metadata
Authors:	Borut Bosančić [1,2], Marija Pecina [2], Nikola Mičić [1,3]
Affiliation:	1. Faculty of Agriculture, University of Banja Luka; 2. Faculty of Agriculture, University of Zagreb; 3. Genetic Resources Institute, University of Banja Luka
Abstract:	<p>In numerous horticultural researches the authors' interest is related to changes in fruit characteristics caused by treatment in one or more genotypes. However, the results are often contradictory. The general significance and the size of the treatment effect can be estimated reliably by meta-analysis using metadata. In horticultural research there are no known developed procedures for the use of meta-analysis which represents a challenge to generate it in order to process the typical multilayer data profile originating from multiple sources and multiple measurements scales. The answer for this typical horticultural metadata setting is application of Main Effect Meta Principal Component Analysis (ME-MetaPCA). ME-MetaPCA represents specific combination of Meta-analysis and Principal Component Analysis adjusted specifically for the use on horticultural data. The goal of this paper is to demonstrate the application and usefulness of this novel technique through simulations and analysis of a typical horticultural data - in apple as a model crop. Data were modeled in terms of plant growth regulator treatment effect researched in several studies on multiple apple varieties. The treatment effect is measured in multiple fruit characteristics on multiple measurement scales. Simulation model is realistically composed of five different studies involving five apple varieties and measuring six fruit characteristics providing total of $5 \times 5 \times 6$, i.e. 150 study level data, which with recorded means, standard deviations and sample size for both treatment and control formed a matrix of 900 data. Besides the standard forest plot as a result of the meta-analysis, the results of the ME-MetaPCA are displayed in the form of biplot in principal components space. This provides straightforward and simple in-one-glance overview of the definite general effects of the studied treatment on i. individual fruit characteristics, ii. studied apple varieties and iii. grouping patterns with grouping rules and exceptions.</p>

Title:	Analysis of influence of seedling type and quality on the outplanting success after disturbances in Norway spruce forest culture
Authors:	Martina Tijardović ¹ , Tomislav Dubravac ¹ , Sanja Perić ¹
Affiliation:	¹ Croatian Forest Research Institute
Abstract:	<p>Norway spruce is the object of numerous discussions because it is susceptible to a number of diverse biotic and abiotic influences, which result with significant economic losses and negative ecological consequences. Areas after disturbances need to be restored, while Norway spruce is often replaced with broadleaves or mixtures of conifer and broadleaved tree species. Research included restoration after disturbances in gaps and on larger area after strong bark beetle attack and windthrow on Bistranska gora locality (Central Croatia). Data collected from 2010. – 2013. (height, basal diameter, quality indices and height increment) were compared using Statistica soft. Repeated measures ANOVA was conducted for different age, quality and type of planting material (European beech seedlings). Replacement by seeding and planting under shelterwood of spruce trees was also included in the research, with the aim of comparison with restoration methods (hemispherical photographs used to document the light availability). The use of the shelterwood method gives the best results for Norway spruce replacement, so it is strongly recommended for one to establish a new forest culture (a mixed one) with a bigger share of broadleaved tree species before disturbance takes place. Nevertheless, the success of ball-rooted seedlings planted after disturbances on the large area showed the best results in terms of height and basal diameter increment, but lowest in terms of quality. Research provides basic knowledge of current restoration needs, risks and silvicultural methods for the Central part of Croatia, as well as a valuable scientific insight into restoration methods on the European level.</p>

Title:	Why (not) study longer? Analysis of duration of undergraduate studies at Faculty of Forestry in Zagreb
Authors:	Mislav Vedriš, Anamarija Jazbec
Affiliation:	Faculty of Forestry, University of Zagreb
Abstract:	<p>Although forestry is often considered as a traditional field of applied science, it resumes its importance in context of increased awareness of climate change and benefits coming from natural ecosystems. Updated and contemporary study programmes are needed to ensure an adequate education on managing forest ecosystems and its products.</p> <p>A better insight in student background and their motivation for studies are welcome to improve the study programmes and also to find the ways how to help students achieve better results. A case study research was carried out to measure success of studies by duration of studies and average grades achieved, with possible causes in the financial support. Three undergraduate study programmes on Zagreb Faculty of forestry were analyzed (Forestry N=94, Urban forestry N=54 and Wood technology N=39), with data acquired from a questionnaire among forestry graduate students on the same faculty academic year 2016/2017. A Factorial ANOVA was performed to test differences between the studies, a Kaplan-Meier estimator for the duration by years and a linear regression for the influence of predictor variables on the duration of study. Although this profession is associated with people from rural areas, about one third of all students comes from the capital city, with significant majority on WoodTech (52%), and 25% and 34% on Forest and UrbanFor, respectively. Students of WoodTech are mostly coming from vocational schools (70%) while Forest and UrbanFor students are mostly from gymnasiums (67% and 83%). For about 91% of students main motivation was a personal sake rather than a family interest (9%). Average scores and the duration of study on three study programmes have not been proven significantly different between the studies, nor between male and female students. A longer time of study is proven to negatively correlate with the average grades on all study programmes: Forest (-0,559), UrbanFor (-0,554) and WoodTech (-0,378). Common predictors for duration (with logarithmic transformation) on all study programmes in regression analysis were the average grade and a personal motivation, both with negative sign, thus leading to a shorter study time. Obtained grant is proven to reduce the duration for WoodTech and Forest, and not for UrbanFor. Type of High school was a good predictor only for WoodTech programme, where a vocational school is shown to increase the duration of study.</p>



BIOSTAT 2017

23rd International Scientific Symposium on Biometrics
 ŠIBENIK, CROATIA, 7 - 10 JUNE 2017

Author	Page	Author	Page
Alekseyenko, Alexander V.	11,12	Mićić, Nikola	26
Armeson, Kent	15	Miller, Cameron	13
Billard, Lynne	10	Neely, Benjamin	13
Bosančić, Borut	26	Obeid, Jihad S.	12
Čaklović, Lavoslav	18	Pecina, Marija	26
Dalbelo Bašić, Bojana	25	Perić, Sanja	27
Dobša, Jasminka	22	Perišić, Ana	20
Drake, Richard	13	Pleše, Borna	17
Dubravec, Tomislav	27	Prot, Franjo	19
Đorđević, Mina	23	Prot, Sara	19
Falak, Hrvoje	20	Rimac, Milan	16
Gusić, Jelena	23	Roganovic, Jelena	16
He, Yongqun "Oliver"	12	Sonicki, Zdenko	16
Hill, Elizabeth G.	13, 15	Stare, Janez	8
Huzurbazar, Snehalata	6	Šimić, Diana	23
Ivanković, Damir	17	Teslak, Krunoslav	24
Jazbec, Anamarija	28	Thompson, Paul A.	14
Kiers, Henk A. L.	22	Tijardović, Martina	27
Kurtanjek, Želimir	21	Urbanc, Jadranka	20
Martin-Harris, Bonnie	15	Vedriš, Mislav	28
Masci, Anna Maria	12	Zheng, Jie	12
Matloff, Norm	14	Žunić, Marijana	24