

**The 47th International
Biometrical Colloquium
47 Międzynarodowe
Colloquium Biometryczne**

Abstracts



Kiry, Poland, 10-14 September 2017

**47 Międzynarodowe Colloquium Biometryczne
Kiry**

**47th International Biometrical Colloquium
Kiry, Poland**

10-14 września 2017

September 10-14, 2017

Organizatorzy konferencji

Organizers

Polskie Towarzystwo Biometryczne
The Polish Biometric Society

Katedra Doświadczalnictwa i Bioinformatyki,
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
Department of Experimental Design and Bioinformatics
Warsaw University of Life Sciences

Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza w Poznaniu
Faculty of Mathematics and Computer Science
Adam Mickiewicz University in Poznań

Katedra Metod Matematycznych i Statystycznych,
Uniwersytet Przyrodniczy w Poznaniu
Department of Mathematical and Statistical Methods
Poznań University of Life Sciences

Komitet naukowy

The Scientific Committee

Ewa Bakinowska (Politechnika Poznańska, Polska)

Ivette Gomes (Faculdade de Ciências, Universidade de Lisboa, Portugal)

Dariusz Gozdowski (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Polska)

Małgorzata Graczyk (Uniwersytet Przyrodniczy w Poznaniu, Polska)

Zofia Hanusz (Uniwersytet Przyrodniczy w Lublinie, Polska)

Monika Jakubus (Uniwersytet Przyrodniczy w Poznaniu, Polska)

Zygmunt Kaczmarek (Instytut Genetyki Roślin PAN w Poznaniu, Polska)

Marcin Kozak (Wyższa Szkoła Informatyki i Zarządzania w Rzeszowie, Polska)

Mirosław Krzyśko (Uniwersytet im. A. Mickiewicza w Poznaniu, Polska)

Izabela Kuna-Broniowska (Uniwersytet Przyrodniczy w Lublinie, Polska)

Wiesław Mądry (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Polska)

Iwona Mejza (Uniwersytet Przyrodniczy w Poznaniu, Polska)

Stanisław Mejza (Uniwersytet Przyrodniczy w Poznaniu, Polska)

Manuela Neves (ISA, Universidade Técnica de Lisboa, Portugal)

Amilcar Oliveira (Universidade Aberta, Lizbona, Portugal)

Teresa Oliveira (Universidade Aberta, Lizbona, Portugal)

Dulce Pereira (Universidade de Évora, Portugal)

Paulo Canas Rodrigues (FCT, Universidade Nova de Lisboa, Portugal)

Leszek Sieczko (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Polska)

Ewa Skotarczak (Uniwersytet Przyrodniczy w Poznaniu, Polska)

Marcin Studnicki (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Polska)

Anna Szczepańska-Álvarez (Uniwersytet Przyrodniczy w Poznaniu, Polska)

Mirosława Wesołowska-Janczarek (Uniwersytet Przyrodniczy w Lublinie, Polska)

Waldemar Wołyński (Uniwersytet im. A. Mickiewicza w Poznaniu, Polska)

Agnieszka Wnuk (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Polska)

Elżbieta Wójcik-Gront (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Polska)

Komitet organizacyjny
Organizing Committee

dr hab. Dariusz Gozdowski - przewodniczący

dr Marcin Studnicki - sekretarz

dr Ewa Bakinowska

dr hab. Monika Jakubus

dr Leszek Siczko

dr Ewa Skotarczak

dr Anna Szczepańska-Álvarez

dr Agnieszka Wnuk

dr Elżbieta Wójcik-Gront

Streszczenia referatów

Multivariate analysis of texture properties of butternut squash in thermal treatment

DARIUSZ ANDREJKO¹, ZOFIA HANUSZ², KAMILA KLIMEK², BEATA ŚLASKA-GRZYWNA¹

¹Department of Biological Bases of Food and Feed Technologies
University of Live Sciences in Lublin, Głęboka 28, 20-612 Lublin

²Department of Applied Mathematics and Computer Science
University of Live Sciences in Lublin, Głęboka 28, 20-612 Lublin

In the paper the changes of texture properties of butternut squash caused by thermal treatment in a convection steam oven with different parameters of the processes. Texture properties of pumpkin pulp were measured by five features, namely, hardness, chewiness, gumminess, springiness and cohesiveness. In the paper Ślaska-Grzywna et al. (2016), statistical analysis of each feature separately was presented. In this paper, to discover impact of three factors of thermal treatment: temperature, steam and time, each at different levels, on five texture features under consideration simultaneously, multivariate approach has been applied. Namely, the study includes multivariate analysis of variance MANOVA and chosen multivariate techniques such as: canonical correlation, principal components, factor, cluster and discriminant analyses (Khattree and Naik, 1999, 2000). All calculations were done using SAS Enterprise Guide and SAS 9.3.

References

- Ślaska-Grzywna B., Blicharz-Kania A., Sagan A., Nadulski R., Hanusz Z., Andrejko D., Szmigielski M. (2016). *Changes in the texture of butternut squash following thermal treatment*. Italian Journal of Food Science 28 (1), 1-8.
- Khattree R., Naik N. (1999). *Applied Multivariate Statistics with SAS Software*. Cary, NC: SAS Institute Inc.
- Khattree R., Naik N. (2000). *Multivariate Data Reduction and Discrimination with SAS Software*. Cary, NC: SAS Institute Inc.

Geostatistical Methods in R

URSZULA BRONOWICKA-MIELNICZUK

Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Poland

urszula.bronowicka@up.lublin.pl

Geostatistics comprises a wide range of statistical techniques that are adjusted to spatial data. In recent years, we have noticed a significant increase in interest in the use of geostatistical methods in agriculture, ecology and environmental sciences. In this study we present a brief introduction to geostatistical data analysis with the R packages *gstat* and *geoR* used in conjunction with the package *sp*.

Keywords: geostatistics, R, spatial prediction, spatial data analysis, variogram modelling, kriging

References

- Bivand R, Pebesma E.J., Gómez-Rubio V. 2008. *Applied Spatial Data Analysis with R*. Springer, New York.
- Gräler B., Pebesma E.J, Heuvelink G. 2016. *Spatio-Temporal Interpolation using gstat*. *The R Journal* 8(1), 204-218.
- Hengl T. 2009. *A practical guide to geostatistical mapping of environmental variables*. 2nd. ed., University of Amsterdam.
- Pebesma, E.J., 2004. *Multivariable geostatistics in S: the gstat package*. *Computers & Geosciences*, 30: 683-691.
- Pebesma, E.J., Bivand R.S., 2005. *Classes and methods for spatial data in R*. *R News* 5 (2). <https://cran.r-project.org/doc/Rnews/>.
- Ribeiro P.J. Jr, Diggle P.J. 2016. *geoR: Analysis of Geostatistical Data*. R. package version 1.7-5.2. <https://CRAN.R-project.org/package=geoR>.
- Wackernagel, H. 1998. *Multivariate Geostatistics. An introduction with applications*, 2nd ed., Springer, Berlin.

**On a new approach to the analysis of variance
for experiments with orthogonal block structure**

TADEUSZ CALIŃSKI, IDZI SIATKOWSKI*

Department of Mathematical and Statistical Methods,
Poznań University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland

*idzi@up.poznan.pl

Estimation and hypothesis testing procedures can be simplified if the experiment has the orthogonal block structure, recalled here.

Special attention is given to the randomization-derived models for analyzing experiments in block and nested block designs.

A data transformation is suggested to make the analysis of variance more simple, particularly if the main interest is in treatment comparisons.

Thus, the analysis of variance based on the randomization-derived model is considered for some experiments with the orthogonal block structure.

Keywords: analysis of variance, block designs, estimation, hypothesis testing, orthogonal block structure, randomization-derived model

D-optimal weighing designs with negative correlations of errors: new classes

BRONISŁAW CERANKA, MAŁGORZATA GRACZYK

Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

In this paper, the properties of the regular D-optimal chemical balance weighing design are considered. We study this design under assumption that the measurements errors are equally negative correlated they have the same variances. Here we investigate the issues regard to the existence conditions of regular D-optimal design. We present the relations between the parameters of such design and construction methods.

Keywords: balanced bipartite weighing design, chemical balance weighing design, D-optimality, ternary balanced block design

References

- Ceranka B., Graczyk M. 2014. *Construction of the regular D-optimal weighing designs with non-negative correlated errors*. Colloquium Biometricum 44: 43–56.
- Ceranka B., Graczyk M. 2014. *Regular D-optimal weighing designs with negative correlated errors: construction*. Colloquium Biometricum 44: 57–68.
- Jacroux M., Wong C.S., Masaro J.C. 1983. *On the optimality of chemical balance weighing design*. Journal of Statistical Planning and Inference 8: 213–240.
- Katulska K., Smaga Ł. 2013. *A note on D-optimal chemical balance weighing designs and their applications*. Colloquium Biometricum 43: 37–45.
- Masaro J., Wong Ch.S. 2008. *D-optimal designs for correlated random vectors*. Journal of Statistical Planning and Inference 138: 4093–4106.

Minimum Information guidelines for harmonisation of plant phenotyping data

HANNA ĆWIEK-KUPCZYŃSKA^{1*}, WOJCIECH FROHMBERG², PAWEŁ KRAJEWSKI¹

¹Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland

²Poznań University of Technology, Poland

*hcwi@igr.poznan.pl

Plant phenotyping leads to producing data of different types and scales which, when linked to genetic information, help explain biological phenomena. Correct interpretation of the collected observations, and thus replicability, comparability and interoperability of data, is possible on the condition that an adequate set of metadata is provided. Standardisation of data description allows creation of generic repositories, tools and services that facilitate datasets' publication, exchange and reuse [1].

The paper presents the results of research conducted with partners of the project Trans-national Infrastructure for Plant Genomic Science (transPLANT, <http://transplantdb.eu>). Within a package "Community standards for the interoperability of data resources", reporting guidelines for plant phenotyping data, called Minimum Information About a Plant Phenotyping Experiment (MIAPPE), have been proposed [2, 3].

We demonstrate how datasets obtained from plant phenotyping experiments can be described following the MIAPPE recommendations, structured in ISA-Tab format [3, 4], and managed in a data repository taking advantage of the recommended metadata characteristics for browsing, search and analysis. An exemplary application of the approach has been done in SEGENMAS project, where several partners performed various assays and measurements on samples obtained from numerous field and laboratory experiments with a set of lupin (*Lupinus angustifolius*) accessions.

Keywords: experimental data, metadata management, minimum information standards, plant phenotyping.

References

- [1] Wilkinson M.D., et al. 2016. *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data 3: 160018.
- [2] Krajewski P., et al. 2015. *Towards recommendations for metadata and data handling in plant phenotyping*. Journal of Experimental Botany 66 (18): 5417–27.
- [3] Ćwiek-Kupczyńska H., et al. 2016. *Measures for interoperability of phenotypic data: minimum information requirements and formatting*. Plant Methods 12: 44.
- [4] Sansone S.A., et al. 2012. *Toward interoperable bioscience data*. Nature Genetetics 44 (2): 121–6.

Acknowledgements

This research has been supported by the European Union under the FP7 thematic priority Infrastructures, project No. 283496 transPLANT, and by Polish National Centre for Research and Development, project SEGENMAS PBS3/A8/28/2015.

A comparison of kernel density estimates and their applications

EWA DIAKOWSKA, ANDRZEJ MICHALSKI, MACIEJ KARCEWSKI, LESZEK KUCHAR

Department of Mathematics
Wroclaw University of Environmental and Life Sciences
Grunwaldzka 53, 50-357 Wroclaw, Poland

In this article we compare and examine the effectiveness of different kernel density estimates for a some experimental data. For a given random sample X_1, X_2, \dots, X_n we present over a dozen kernel estimators \hat{f}_n of the density function f with Gaussian kernel and with the kernel given by Epanechnikov (Berliner and Devroye, 1994) using four methods: Silverman's rule of thumb, Sheather-Jones method, cross-validation method and the plug-in method (Feluch and Koronacki, 1992, Silverman, 1986, Givens and Hoeting, 2005, Wand and Jones, 1995). For assessing the effectiveness of considered estimates and their similarity we applied a distance measure for measurable and integrable functions (Marczewski and Steinhaus, 1958). All numerical calculations were done for an experimental data recording groundwater level on a melioration facility (cf. Michalski, 2016).

References

- Berliner, A., Devroye, L., 1994. *A comparison of kernel density estimates*. Publications de l'Institut de Statistique de l'Université de Paris, vol. XXXVIII –Fascicule 3, 3-59.
- Feluch, W., Koronacki, J., 1992. *A note on modified cross-validation in density estimation*. Computational Statistics & Data Analysis, 13, 143-151.
- Givens, G.H., Hoeting, J.A., 2005. *Computational Statistics*. New York: Wiley & Sons.
- Marczewski, E., Steinhaus, H., 1958. *On a certain distance of sets and the corresponding distance of functions*. Colloquium Mathematicum 6, 319-327.
- Michalski, A., 2016. *The use of kernel estimators to determine the distribution of groundwater level*, *Meteorology Hydrology and Water Management*. Vol. 4, 1, 41-46, DOI:10.26491/mhwm/62708
- Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.
- Wand, M.P; Jones, M.C., 1995. *Kernel Smoothing*. London: Chapman & Hall.

**The problems of multivariate analyses (PCA and FA)
of categorical - ordinal variables in sociology -
Review of correlations coefficients**

PAVEL FLAK

Biometry Commission of the Slovak Academy of Agricultural Sciences
Hlohovecká 2, 949 41 Lužianky (Nitra), Slovak Republic,
e-mail: flak@vuzv.sk
Private: Tr. A. Hlinku 10, 949 01 Nitra, Slovak Republic
e-mail: flak14@azet.sk

The Principal Component Analysis (PCA) and Factor Analysis (FA) are widely used statistical techniques in the psychological and social sciences. But in these sciences are analysed not only continuous variables but mainly categorical ones (i.e. binary, dichotomous, mainly ordered categories - ordinal variables). In practice most researchers treat ordinal variables with 5 or more categories as continuous. But ordinal variables with many categories are often nonnormal and also with nonhomogeneous in variability. There are various specific statistical approaches of analysis of these data, e.g. by Ordinal Factor Analysis. In present paper is a short review of various types of correlations coefficients, which have important role in sociology and mainly in psychology sciences.

Keywords: biometrics, sociology, psychology, categorical variables, correlation, multivariate statistical analyses.

Variable selection for classification of multivariate functional data

TOMASZ GÓRECKI, MIROSLAW KRZYŚKO*, WALDEMAR WOŁYŃSKI

Faculty of Mathematics and Computer Science
Adam Mickiewicz University, Poznań, Poland

*mkrzysko@amu.edu.pl

New variable selection method is considered in the setting of classification with functional data $\{X(t) = (X_1(t), \dots, X_p(t))', t \in I\}$ (Ramsay and Silverman (2005), Horváth and Kokoszka (2012)). The variable selection is a dimensionality reduction method which leads to replace the whole vector process $X(t)$, with a low-dimensional vector $(X_{(1)}(t), \dots, X_{(d)}(t))'$ still giving a comparable classification error. The various classifiers appropriate for functional data are used. The proposed variable selection method is based on functional distance covariance (Székely and Rizzo (2009, 2012)) and is a modification of the procedure given by Kong et al. (2015). The proposed methodology is illustrated on a real data example.

Keywords: multivariate functional data, variable selection, functional distance covariance, classification

References

- Horváth L., Kokoszka P. 2012. *Inference for Functional Data with Applications*. Springer. New York.
- Kong J., Wang S., Wahba G. 2015. *Using distance covariance for improved variable selection with application to learning genetic risk models*. *Statistics in Medicine* 34: 1708-1720.
- Ramsay J.O., Silverman B.W. 2005. *Functional Data Analysis, Second Edition*. Springer. New York.
- Székely G.J., Rizzo M.L. 2009. *Brownian distance covariance*. *Annals of Applied Statistics* 3(4): 1236-1265.
- Székely G.J., Rizzo M.L. 2012. *On the uniqueness of distance covariance*. *Statistical Probability Letters* 82(12): 2278-2282.

Evaluation of classification of crop fields with maize using satellite data from Sentinel-2

DARIUSZ GOZDOWSKI* , KAROL ZGLEJSZEWSKI

Warsaw University of Life Sciences, Department of Experimental Design and Bioinformatics

*dariusz_gozdowski@sggw.pl

Land cover classification using satellite data is used for many purposes including agricultural policy. It is important e.g. for control of crops declared by farmers for subsidies which receive from EU funds. For such purpose are used satellite imagery of very high resolution (pixel size about 1 m). Because such satellite are quite expensive the alternative can be free available satellite data from e.g. Sentinel-2 (pixel size 10 m).

In year 2017 the study in which multi-temporal (from spring to the end of summer) satellite data from Sentinel-2 were used for classification of selected crops. One of them was maize which is characterized by different pattern of growth than other cereals species cultivated in Poland, i.e. later sowing and harvest date. Multispectral satellite data were used for the analyses. Cluster analysis and logistic regression were applied for multivariate classification. Variables were spectral bands in visible and infra-red range and vegetation indices such as NDVI (normalized difference vegetation index). The examined fields with maize were located in Mazovia region and were divided into training set and validation set to evaluate sensitivity and specificity of classification. The obtained results are promising especially for bigger fields and the classification can be useful at regional scale in north and wester part of Poland where field size is bigger.

Keywords: multivariate classification, satellite data, remote sensing, crops

Acknowledgements

This research has been supported by the European Space Agency, project *FERTISAT - Satellite-based Service for Variable Rate Nitrogen Application in Cereal Production*.

Damage-Robust alpha design

DAVID HAMPEL^{*}, JITKA JANOVÁ

Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, Brno, 61300, Czech Republic

^{*}qqhampel@mendelu.cz

When using alpha-design for plant variety testing under space restrictions, ex post design modifications must be implemented to prevent variety self-proximity on plots and, consequently, to prevent damage-induced loss of experimental information. This is done ad hoc for each experiment; the unsystematic modification is, however, commonly not only unable to resolve all existing proximities, but may introduce secondary undesired proximities. In this presentation, a procedure is developed for the universal construction of modified alpha-design that covers all existing proximity constraints while keeping the efficiency level of the original design. Using extensive real data simulation, we validate the procedure and confirm high damage robustness of the modified designs. The procedure has been implemented as a Matlab function and is available as on-line supplement to the paper Janová and Hampel (2016). The function enables to design the damage-robust experiments automatically using only standard computer equipment.

Keywords: Alpha-design, damage robustness, proximity constraints, simulation, variety proximity

References

- Janová J., Hampel D. 2016. *alfaDRA: A program for automatic elimination of variety self-proximities in alpha-design*. Computers and Electronics in Agriculture 122: 156–160.
- John J.A., Williams E.R. 1998. *t-latinized designs*. Australian & New Zealand Journal of Statistics 40, 111–118.
- Patterson H.D., Williams E.R. 1976. *A new class of resolvable incomplete block designs*. Biometrika 63: 83–92.

Acknowledgements

Supported by the grant No. GA13-25897S of the Czech Science Foundation.

Practical applicability of germination index assessed by logistic models

MONIKA JAKUBUS^{1*}, EWA BAKINOWSKA²

¹Department of Soil Science and Land Protection
Poznan University of Life Sciences, Poland

² Institute of Mathematics, Poznan University of Technology, Poznan, Poland

*e-mail:monja@up.poznan.pl

Sewage sludge management is a major challenge in environmental protection. Composting is an organic waste treatment method that is cost effective and leads to resource recovery. Composting is considered an environmentally and agriculturally friendly method of sewage sludge utilisation. The objective of this study was to evaluate maturity of three composts prepared on the basis of sewage sludge mixed with structure-forming waste materials such as pine bark, sawdust and wheat straw. The germination index (GI) was used to assess the maturity and phytotoxicity of composts at particular composting stages (initial, mesophilic, thermophilic, cooling, maturation). Cress seeds were used to determine the germination index (GI). The logistic model, which belongs to a broad class of generalized linear models, was used to analyse experimental data. Using this model the interesting probabilities (from the point of view of the experimenter) for the occurrence of a specific root length were determined. In addition, a model was constructed providing a dependence of probability on temperature.

This work indicates a marked dependence between root length produced by cress seeds and the temperature of the composting process, which was closely related to the GI values. The longest plant roots, similarly as the highest GI values, were found at the lower temperature, which took place at the beginning and at the end of the composting process. Our findings suggest that the practical applicability of GI in the evaluation of compost maturity is limited. Additionally, the role of additional wastes being structure-forming agents in composted mixtures with sewage sludge was stressed as a sorption matrix for harmful substances released from sewage sludge.

Keywords: germination index, cress seeds, composting process, sewage sludge composts, logistic model

Changes of various sulphur fractions during co-composting of pine bark with plant material

MONIKA JAKUBUS^{1*}, MAŁGORZATA GRACZYK²

¹Department of Soil Science and Land Protection
Poznan University of Life Sciences, Poland

²Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

*monja@up.poznan.pl

Composting pine bark alone and with addition is an interesting alternative to recycling waste as a compost. Such produced composts are valuable sources of nutrients and organic matter. The study has been focused on the exploration of various sulphur fractions (total, plant available, easily mineralizable organic and residual) in four composts during progressive composting process. Experimental composts were prepared by using scots pine (*Pinus silvestris* L.) bark and chopped plant material (PM) according to the scheme: C1-pine bark, C2-pine bark mixed with urea (dose of urea was applied at amount equivalent to 1kg N per 1 m³ of pine bark), C3-pine bark mixed with PM (0.5 Mg of PM per 1 m³ of pine bark), C4-pine bark mixed with PM (3.5 Mg of PM per 1 m³ of pine bark). For sulphur extraction two single procedures were used with solution of CH₃COOH and KCl. The period of composting process lasted 203 days and comprised of 6th stages. With regards to the aim of study the evaluation of quantitative changes of sulphur was analysed by different statistical tools as analysis variance and principal component analysis. To examine the influence of composts and stages, the analysis of variance in the model of two-way cross classification with interaction was considered. It was found, that compost prepared on the basis of pine bark and plant material (C4) showed the highest amounts of sulphur fraction and their changes were significant during the composting process. The use of PCA to summarize the influence of composts and stages has been found to be valuable and not routine in such issues. The data of PCA proved that with regard to the plant available sulphur and easily mineralizable organic sulphur the composting process could be shortened to 80th days without decreasing the quality of composts. The content of total and residual sulphur underwent to similar pattern of variation. The amounts of sulphur extracted with CH₃COOH and KCl as well as and their changes observed during the composting process were comparable. However the solution of KCl may be pointed as more sensitive extractor of sulphur in composts.

Keywords: pine bark, various S fractions, single extractors, composting, PCA

Comparison of binomial proportions: new test

STANISŁAW JAWORSKI, WOJCIECH ZIELIŃSKI*

Faculty of Applied Informatics and Mathematics
Warsaw University of Life Sciences, Poland

*wojtek.zielinski@statystyka.info

In the problem of comparison of two probabilities of success the most widely used is approximate test based on de Moivre - Laplace theorem. In the paper a test based on likelihood ratio is proposed. Those tests are compared due to probability of an error of the first kind. A medical example is presented.

Keywords: binomial proportions, comparison of probabilities of success

A note about the moments of the product of random variables

ANDRZEJ KORNACKI

Department of Applied Mathematics and Computer Science

University of Life Sciences in Lublin

ul. Akademicka 12 20-950

Andrzej.Kornacki@up.lublin.pl

In many areas of the natural sciences eg (physics, economics, finance) we meet quantities which are products of other quantities In statistics, they correspond to estimators that are products of other estimators. Hence, the question arises how to express moments of a product by moments of factors. In this paper we present such formulas . In the general case, the formulas are very complex and unreadable. Taking into account practical applications, the moments of orders $r = 2,3,4$. are important They are the starting point for the calculation of variance, asymmetry coefficient or kurtosis.(Fisz 1967, Rao 2002) For such situations, the article gives both accurate and approximate formulas. The efficiency of approximation was also estimated for variance The results for $r = 3$ and 4 are generalizations of the formulas for variance given by Goodman (Goodman 1960, 1962).

Keywords: central moment of random variable, independence of random variables, coefficient of assymetry, kurtosis, coefficient of variation, approximate formula.

References

Fisz, M 1967. Rachunek prawdopodobieństwa i statystyka matematyczna. Warszawa PWN.

Goodman L.A 1960. *On the exact variance of products* JASA, 55,708-713.

Goodman L.A 1962. *The variance of product of random variables* JASA, 57,54-60.

Rao C.R 2002. Linear Statistical Inference and Its Applications. John Wiley & Sons New York.

Designing factorial experiments in plant protection research

MARIA KOZŁOWSKA

Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

markoz@up.poznan.pl

The specific character of research on plant protection implies the necessity to studies on planning and analysis of such experiments (e.g. Kozłowska 2014, Kozłowska et al. 2014). Plant protection experiments are designed on heterogeneous experimental material, for example, in the case of non-uniform or particularly low levels of disease or pest infection. There are frequently factorial or near-factorial experiments. In such experiments the importance of interaction and hidden replication are emphasized. The designing of plant protection experiments is complicated. Block design with nested rows and columns is frequently used (e.g. Kozłowska 2001). A design is said to have nested rows and columns if the set of experimental units is partitioned into blocks and each block is further partitioned into rows and columns. Thus it is reasonable to seek a design that can withstand the loss of blocks. Several authors have investigated conditions for robustness of some block designs (e.g. Godolphin and Warren 2011, Godolphin and Godolphin 2015). The robustness of block design with nested rows and columns in the face of loss of whole blocks are presented. Examples of block designs with nested rows and columns applied to the plant protection experiments mentioned above are described.

Keywords: block design with nested rows and columns, plant protection experiment, designing experiment

References

- Godolphin J. D. Godolphin E. J. 2015. *The robustness of resolvable block designs against the loss of whole blocks or replicates*. Journal of Statistical Planning and Inference 163: 34-42.
- Godolphin J. D. Warren H. R. 2011. *Improved conditions for the robustness of binary block designs against the loss of whole blocks*. Journal of Statistical Planning and Inference 141: 3498-3505.
- Kozłowska M. 2001. Planowanie doświadczeń z zakresu ochrony roślin w układach blokowych z zagnieżdżonymi wierszami i kolumnami. Roczniki AR w Poznaniu 313, Rozprawy naukowe, Poznań.
- Kozłowska M. 2014. Przewodnik po dobrej praktyce eksperymentalnej. PTB, PRODRUK, Poznań.
- Kozłowska M., Jaskulska M., Łacka A., Kozłowski R. J. 2014. *Analysis of studies of the effectiveness of a biological method of protection for organic crops*. Biometrical Letters 51(1): 45-56.

Genome-metry: statistical methods in genome-level data analysis

PAWEŁ KRAJEWSKI^{1*}, ANETA SAWIKOWSKA^{1,2}

¹Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland

²Department of Mathematical and Statistical Methods

Poznań University of Life Sciences, Poland

*pkra@igr.poznan.pl

An increasing number of research questions in biology is today answered by high-throughput sequencing of DNA or RNA. The experiments utilizing those techniques grow in size and complexity, and the next-generation sequencing (NGS) data volume expands. In consequence, the field of investigations that has been linked mainly to bioinformatic methods is increasingly using the biometrical methods rooted in classical statistical methodology. The contradicting requirements that we are faced with during the data analysis are a relatively small number of biological replications and a high number of features. We will present some approaches that became standards in the analysis of genomic data and also new approaches that we used in the analysis of real experimental results in collaboration with plant biologists. The presentation will cover a range of applications of NGS protocols.

Keywords: next generation sequencing, data analysis, genomic features

Acknowledgements

This research is supported through the ERA-CAPS project 'FlowPlast' by a grant from the Polish National Centre for Research and Development (ERA-CAPS-1/1/2014).

Morphological features selection for the chosen tree taxa

AGNIESZKA KUBIK-KOMAR^{1*}, ELŻBIETA KUBERA¹,
KRYSTYNA PIOTROWSKA-WERYSZKO²

¹Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Poland

²Department of Botany
University of Life Sciences in Lublin, Poland

*agnieszka.kubik@up.lublin.pl

The basis of aerobiological studies is the monitoring of pollen concentration in the air and the term of pollen season. This task is performed by appropriately trained staff and is difficult and time consuming.

The goal of this research is to select the morphological characteristics of the grains that are the most discriminative for distinguishing between birch, hazel and alder taxa and are easy to determine automatically from the microscope images. This selection is based on the split attributes of the classification trees J4.8 built for different subsets of features. The most discriminative among the studied 13 morphological characteristics are: the number of pores, maximal axis, minimal axis, axes difference, maximal oncus width, the number of sided pores. The classification result of the tree based on this subset is better than the one built on the whole feature set. Therefore, the attributes selection before tree building is recommended before tree building, because of the possibility of classifier overfitting.

The classification results for the features easiest to obtain from the image, i. e. maximal axis, minimal axis, axes difference, and the number of sided pores, are only 2.67 pp lower than those obtained for the complete set, but up to 4 pp lower than on the results obtained for the selected, most discriminating attributes only .

Keywords: pollen grains, morphological features, classification tree

References

- Breiman L., Friedman J.H., Olshen R.A., Stone C. J. 1984. *Classification and Regression Trees* (2nd Ed.). Pacific Grove, CA; Wadsworth
- Kubik-Komar A, Kubera E, Piotrowska-Weryszko K, Weryszko-Chmielewska E. 2015. *Application of decision tree algorithms for discriminating among woody plant taxa based on the pollen season characteristics*. Archives of Biological Sciences, 67 (4), 1127-1135
- Piotrowska, K. , Kubik-Komar, A. 2012. *The effect of meteorological factors on airborne Betula pollen concentrations in Lublin (Poland)*. Aerobiologia, 28 (4), 467-479

Modeling of the strength of mineral fertilizer granules

NORBERT LESZCZYŃSKI¹, MAŁGORZATA SZCZEPANIK^{2*}, MILAN KOSZEL³

¹Department of Horticultural and Forest Machinery
University of Life Sciences in Lublin, Poland

²Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Poland

³Department of Management and Machine Exploitation
University of Life Sciences in Lublin, Poland

*malgorzata.szczepanik@up.lublin.pl

The experiments were aimed at determining the relationship between the granule-breaking force and the speed of compression. Five size fractions of granules of three mineral fertilizers were subjected to strength tests using six compression speeds. It was found that the force that broke fertilizer granules depends on the make of the fertilizer, on its granularity and the logarithm of speed at which granules are compressed. The changes in the breaking force were described by regression equations which were subsequently compared with each other.

Keywords: strength of particles, mineral fertilizers, regression lines, hypothesis testing

References

- Bonakdar T., Ghadiri M., Ahmadian H., Bach P. 2015. *Impact Strength Distribution of Placebo Enzyme Granules*. Powder Technology, 285, 68–73.
- Ghasemi Y., Kianmehr M. H. 2016. *The effect of filling of the drum on physical properties of granulated multicomponent fertilizer*. Physicochemical Problems of Mineral Processing 52(2), 2016, 575–587.
- Jamshidian M., Liu W., Zhang Y., Jamshidian F. 2005. *SimReg: A software including some new developments in multiple comparison and simultaneous confidence bands for linear regression models*. Journal of Statistical Software 12(2), 1–23.
- Leszczyński N., Przystupa W., Nowak J., Tatarczak J., Kowalczyk J., Rusek P. 2016. *Studies on compression strength of superphosphate and urea granules*. Przemysł Chemiczny, 95(1): 121–124.
- Rozenblat Y., Portnikov D., Levy A., Kalman H., Aman S., Tomas J. 2011. *Strength distribution of particles under compression*. Powder Technology, 208, 215–224.
- Salman A.D., Fu J., Gorham D.A., Hounslow M.J. 2003. *Impact breakage of fertiliser granules*. Powder Technology, 130, 359–366.
- Wojnar J., Zieliński W. 2004. *Multiple comparisons of slopes of regression lines*. Biometrical Letters 41(1), 25–35.

Uniform smoothness and uniform rotundity in Banach spaces

JOANNA MARKOWICZ*

¹Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin, Poland

*joanna.markowicz@up.lublin.pl

The aim of this paper is to present some geometric properties of Banach spaces, i.e. rotundity and smoothness. We shall start with the definition of smooth spaces and give the fundamental properties of such spaces with some examples. We shall define the function called the modulus of smoothness of a normed space X . We shall show the basic properties of that function. We shall define uniformly smooth spaces and give the relationship between those spaces and the modulus of smoothness.

In the next part of the paper we shall introduce the rotund spaces (strictly convex/strictly norms spaces) adding some examples. For a normed space X we shall define the function called the modulus of rotundity (the modulus of convexity). Furthermore, we shall give the definition and the characterization of the uniformly convex spaces.

The paper finishes with the presentation of a duality between uniform smoothness and uniform rotundity. We shall examine that relationship.

Keywords: geometry of Banach spaces, smooth spaces, uniform smoothness, modulus of smoothness, rotund spaces, uniform convexity, modulus of convexity

References

- Chidume C., 2009. *Geometric properties of Banach Spaces and Nonlinear Iterations, Lecture Notes in Mathematics*. Springer-Verlag London.
- Convay J. B., 1990. *A Course in Functional Analysis*. Springer-Verlag New York.
- Goebel K. and Kirk. W. A., 1999. *Zagadnienia metrycznej teorii punktów stałych*. Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin.
- Kirk W. A. and Sims B. (eds), 2001. *Handbook of Metric Fixed Point Theory*. Kluwer Academic Publishers: 93-132.
- Prus S., 1999. *On infinite dimensional uniform smoothness of Banach spaces*. Comment. Math. Univ. Carolin. 40,1 (1999) 97-105.

Suitability of selected blackcurrant (*Ribes nigrum* L.) parental genotypes for applied breeding of dessert cultivars

AGNIESZKA MASNY*, STANISŁAW PLUTA, ŁUKASZ SELIGA

Department of Breeding of Horticultural Crops
Research Institute of Horticulture
96-100 Skierniewice

*Agnieszka.Masny@inhort.pl

Studies were conducted in 2012-2014 at the Research Institute of Horticulture in Skierniewice, Poland. The aim of the research was to assess the breeding value, based on the effects of general and specific combining abilities (GCA and SCA) of six dessert parental forms of blackcurrant for selected traits. The plant material were seedlings of F₁ generation obtained by crossing of 6 blackcurrant genotypes: 'Big Ben', 'Bona', 'Ceres', 'Sofievskaya', 'Vernisazh' and clone D13B/11. The diallel cross mating design (Griffing's Method III) was used for hybridization of parental forms; in total 30 cross combinations were performed.

The field experiment was established at the Experimental Orchard at Dąbrowice (near Skierniewice), Central Poland. The random block design was used, seedlings were planted in a density of 3.0 x 0.50 m in 3 replications, with 15 plants per plot. Each hybrid family (cross combination) was represented by 45 seedlings and total 1350 seedlings were planted in the field. The seedlings were individually evaluated for plant growth, bush habit, fruit yield and weight of 100 berries, fruit quality (content of extract and vitamin C), plant susceptibility to powdery mildew (*Podosphaera mors-uvae*), anthracnose (*Drepanopeziza ribis*), and white pine blister rust (*Cronartium ribicola*).

Significant and positive GCA effects were obtained for the following cultivars and traits: 'Sofijejskaya' for improving plant growth, extract content in fruit and plant resistance to powdery mildew and anthracnose. 'Vernisazh' was a donor of strong plant growth, high fruit yield, high content of extract and vitamin C in fruits and low plant susceptibility to powdery mildew. 'Big Ben' gave the offspring such attributes as strong plant growth, large fruit and low plant susceptibility to powdery mildew. The breeding clone D13B/11 contributed to the increase of fruit weight and plant resistance to powdery mildew of seedlings. Cultivars 'Bona' and 'Ceres' showed the least suitability for breeding of dessert blackcurrants. For these parental forms the most negative GCA effects were estimated, including plant growth, fruit yield and plant susceptibility to powdery mildew.

Significantly positive values of SCA effects were estimated for the following hybrid families: 'Bona' x 'Big Ben', 'Ceres' x 'Vernisazh', D13B/11 x 'Vernisazh' - high content of extract and vitamin C in fruit; 'Bona' x D13B/11 - strong growth, high content of vitamin C in fruit and low plant susceptibility to powdery mildew; 'Bona' x 'Vernisazh' - high content of vitamin C in fruits and low plant susceptibility to powdery mildew; 'Big Ben' x 'Ceres' - strong growth and low plant susceptibility to powdery mildew; 'Big Ben' x 'Sofijewskaja' - large fruit and high content of vitamin C in fruit.

Keywords: breeding value, GCA, SCA, trait, hybrid family.

Acknowledgements

This work was performed in the frame of multiannual programme "Actions to improve the competitiveness and innovation in the horticultural sector with regard to quality and food safety and environmental protection", financed by the Polish Ministry of Agriculture and Rural Development.

A note on a statistical analysis of the chlorophyll content in different stages of the maize variety ‘Anjou 258’

IWONA MEJZA¹, JAN BOCIANOWSKI¹, KATARZYNA AMBROŻY-DERĘGOWSKA¹,
KAMILA NOWOSAD², STANISŁAW MEJZA¹, PIOTR SZULC³, MAŁGORZATA JAGŁA³

¹Department of Mathematical and Statistical Methods,
Poznań University of Life Sciences, Poland

²Department of Genetics, Plant Breeding and Seed Production,
Wrocław University of Environmental and Life Sciences, Poland

³Department of Agronomy, Poznań University of Life Sciences, Poland

The paper deals with an additional particular analysis of the four-year (2005-2008) study on the chlorophyll (a , b , $a+b$) content of the maize variety ‘Anjou 258’. The field trial was conducted every year in a split-plot design at the Agricultural Experimental Station in Swadzim (Poland). This inference is based on three-way analysis of variance technique supported by the theory of contrasts. Particular attention is paid to estimation and testing some comparisons among treatment combination effects connected with six doses of urea $\text{CO}(\text{NH}_2)_2$ and three doses of elemental sulphur through examined series of years. During the years considered the results indicate a statistically significant influence of nitrogen effects on the mean chlorophyll (a , b , $a+b$) content only for ear blooming stages of this maize. It was also found a significant interaction between nitrogen and the years for these traits and three-way interaction of nitrogen with the years and sulphur (apart of one trait).

Keywords: contrasts, Dunnett test, split-split-plot design, the maize variety ‘Anjou 258’

References

- Arnon D.I. 1949. *Copper enzymem in isolated chloroplasts. Polyphenoloxidase in Beta vulgaris*. Plant Physiology 24, 1-15.
- Pearce S.C., Caliński, T., Marshall T.F. de C. 1974. *The basic contrasts of an experimental design with special reference to the analysis of data*. Biometrika 54, 449-460.
- Szulc P., Ambroży-Deręgowska K., Mejza I., Nowosad K., Mejza S., Bocianowski J., Jagła M. 2016. *Linear statistical inference on the chlorophyll content in different stages of the maize variety ‘Anjou 258’*. Colloquium Biometricum 46, 83-98.
- Szulc P., Bocianowski J., Rybus-Zajac M. 2012. *The effect of soil supplementation with nitrogen and elemental sulphur on chlorophyll content and grain yield of maize (Zea mays L.)*. Žemdirbystė-Agriculture 99(3), 247-254.

Information value of indicators species for estimation of the Macrophyte Index for Rivers

KRZYSZTOF MOLINSKI¹, ANNA BUDKA^{1*}, KRZYSZTOF SZOSZKIEWICZ²

¹Poznan University of Life Sciences, Department of Mathematical and Statistical Methods

²Poznan University of Life Sciences, Department of Ecology and Environment Protection
Poznan University of Life Sciences, Poland

*budka@up.poznan.pl

Water is a fundamental condition for human's existence. Exponential development of civilization triggers of a great pressure on the environment. This is a reason why any activity that stops water degradation despite of classes of purity becomes very important. Developed methods for assessments and monitoring of aquatic ecosystems to detect the effects of human activity an environment and decrease the process of their degradation. The Macrophyte River Index (MIR) is an biological indicator of flowing water quality based on occurrences of chosen macrophytes. Inidicator species and MIR calculation are elements of Macrophyte River Test Method.

The study focused on a problem of uncertainty in the measurement, which is a result of incomplete pooled data of indicator species found in tested habitat and used in assessment of Macrophyte River Index. Not taking into account this lacks may result in errors in environmental decisions. The analyzes were carried out on the basis of research conducted in 2008-13 on rivers belonging to one type of abiotic - medium lowland rivers with sandy bottom substrate, on 20 places in all classes of purity in range (I-V). Macrofite species have been identified based on estimated entropy and natural resources such as average environmental trophism, the measurement of ecological tolerance of the species and their coverage. A list of species that have the greatest impact on MIR's value was indicated. This will allow to determine Macrofite Index River more accurate and to reduce redundant workload.

Keywords: river, information value, macrophyte

Analiza pozostałości substancji czynnych środków ochrony roślin w płodach rolnych

KINGA NORAS^{1*}, LESZEK SIECZKO¹, ADAM NORAS¹

¹ Department of Experimental Design and Bioinformatics, Faculty of Agriculture and
Biology, Warsaw University of Life Sciences
Poland

* kinga_noras@sggw.pl

Wraz ze wzrostem intensyfikacji rolnictwa wzrasta ilość stosowanych środków ochrony roślin, a wraz z nią ryzyko nieprawidłowego ich stosowania. Nieprawidłowe stosowanie środków ochrony roślin, niezgodnie z zasadami dobrej praktyki rolniczej, przyczynia się do powstawania pozostałości substancji czynnych zarówno w płodach rolnych jak i w glebie, a także przedostawania się ich do wód gruntowych czy powierzchniowych. W pracy podjęto próbę odpowiedzi na kilka strategicznych pytań odnośnie prawidłowości stosowania środków ochrony roślin, w tym stosowania integrowanych metod ochrony, wpływu zużycia środków ochrony roślin na poziom pozostałości środków ochrony roślin w płodach rolnych. Celem pracy jest opracowanie modelu statystycznego z odpowiednimi założeniami, który będzie adekwatny do rozpatrywanych danych. Etap opracowania modelu będzie uwzględniał ocenę różnych modeli, gdzie zdefiniowane będą różne struktury macierzy kowariancji. Do oszacowania efektów wykorzystana zostanie metoda największej wiarygodności z restrykcją REML (ang. Residual Maximum Likelihood). Po wyborze najlepszego modelu, czyli tego, który najlepiej opisuje analizowane dane nastąpi etap wnioskowania i odpowiedzi na pytania.

Keywords: pozostałości środków ochrony roślin, integrowana ochrona roślin, REML

Evaluation of effectiveness of different analysis schemes of dominant farming systems based on aggregated data from National Agricultural Census

MARCIN OLLIK

¹Department of Experimental Design and Biostatistics
Warsaw University of Life Sciences SGGW

marcin_ollik@sggw.pl

The fundamental problem with classification based on a large data set is a number of variables. Two assumptions should be met: the number of cases have to distinctly higher than number of variables and cluster analysis should not be conducted on too many variables. This difficulty should be solved during the analysis of dominant farming systems in Poland based on National Agricultural Census 2010 data. The aim of this work was comparison of methods to reduce the number of variables.

For these work was performed analysis on county level (LAU1, NUTS4). Based on NAC data supplemented by data about land use and population density were prepared 109 variables belonged to 8 groups: farm size structure, agricultural land use, crop structure, animal production, fertilization, mechanization, income sources and county land use.

Three schemes of number of variables reduction was compared:

- The first was Principal Component Analysis performed on all 109 variables
- The second was PCA performed on variables in 8 group separately.
- The third was simply chose 23 original variables (2 – 3 from each group) best characterized by agriculture in Poland

After number of variables reduction counties was grouped by cluster analysis (Euclidean distance, Ward method) by 10 groups and mapped. Results of all three methods was rated by effectiveness of separating the variability of counties, spatial consistency of clusters and correspondence with Polish geography.

All three methods have proven effective in describing agricultural diversification in Poland. However, the variable selection method was distinctly weaker than both PCA-based.

Keywords: farming system, National Agricultural Census 2010, PCA, cluster analysis

Prediction accuracy vs describing ability for factor-analytic linear mixed models for winter wheat multi-environmental trials

JAKUB PADEREWSKI¹, MARCIN STUDNICKI¹, HANS PETER PIEPHO²
ELŻBIETA WÓJCIK-GRONT¹

¹ Dept. of Experimental Design and Bioinformatics, Warsaw University of Life Sciences
² Biostatistics Unit, Institute of Crop Science, University of Hohenheim

Some of the most popular statistics are the post-hoc significance tests. The describing ability with the low numbers of parameters is the key for significance. The mixed models are the most often used methods in agronomy trials. This group of methods assumes different structures for covariance between random factors. Instead describing ability we can consider the prediction ability of the models that is the ability of appropriate prediction not knowing (during the estimation process) cases. The aim of this research was to compare the describing ability and prediction ability of different mixed models for post-registration winter wheat multi-environmental trials.

Required sample sizes for multiple tests

DARIUSZ PARYS*

University of Lodz
Department of Statistical Methods,

In multiple hypothesis testing connecting with multiple comparisons we must consider the control of measure of error rates like FWER (Familywise Error Rate), FDR (False Discovery Rate), pFDR and others. In this paper we present the required sample sizes formulas for multiple tests in the context of controlling the error rates.

The large increase in the number of comparisons often only requires a small increase in the sample size.

Keywords: multiple comparisons, multiple tests, sample size, error rates

Comments on planning of block (variety) trials

WIESŁAW PILARCZYK

Department of Mathematical and Statistical Methods
Poznań University of Life Sciences, Poland

wpilar@up.poznan.pl

There are many factors influencing the precision of treatment comparisons in field trials. Some of these factors (plot size and shape, block size, number of replicates) are related to the extent and character of soil variability. The other factors are related to the mathematical model of observation and methods of statistical analysis of trial results (analysis of variance, analysis of covariance, methods that incorporate the spatial relationship among data). Some additional factors can also influence the trial precision. Among them neighboring influences can be mentioned and border effects as well. In a paper all mentioned factors are shortly discussed and illustrated using extensive trial data from Polish (and Czech) variety testing. Some methods of assessing trial correctness are also discussed

Keywords: block trial, method of statistical analysis, trial precision

References

- Pilarczyk W. 1991. *The efficiency of α -designs in Polish variety testing field trials*, Plant Varieties and Seeds 4, 37-41
- Pilarczyk W. 2009. *The extent and prevailing shape of spatial relationships in Polish variety testing trials on wheat*, Plant Breeding 128, 411-415.
- Pilarczyk W. 1990. *Skuteczność różnych metod analizy jednoczynnikowych doświadczeń blokowych*, Wiadomości Odmianoznawcze 2/39, 1-101.
- Pilarczyk W. 1990. *Sąsiedzki wpływ wysokości roślin na plonowanie odmian zbóż*, Wiadomości Odmianoznawcze 3/40, 35-42.
- Pilarczyk W., Hartmann J. 2004. *Extent of spatial dependence in Czech and Polish cereal variety trials*, Datastat'2003, Folia Fac. Sci. Nat. Univ. Masaryk Brunensis, Mathematica 15, 299-306
- Williams E. 1986. *A neighbour model for field experiments*, Biometrika 73(2), 279-287.

A data analysis protocol for monitoring metabolomic changes in barley under drought stress

ANETA SAWIKOWSKA^{1,2,*}, PAWEŁ KRAJEWSKI²,
ANNA PIASECKA^{2,3}, PIOTR KACHLICKI²

¹Department of Mathematical and Statistical Methods,
Poznań University of Life Sciences, Poland

²Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland

³Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

* asaw@igr.poznan.pl

A data analysis protocol, from raw observations to the statistical descriptors, for large chromatographic data sets acquired in multifactorial experiments is presented. We present an application of the algorithms to data obtained in a study of the reaction of barley to drought.

Data pre-processing consists of several integrated stages, some performed using publicly available software (R), some with our own scripts written for known algorithms, such as chromatogram alignment by correlation optimized warping, and others with our own, new algorithms.

The statistical approach is based on the analysis of variance performed with the help of the restricted maximum likelihood (REML) numerical procedures in Genstat package. The analysis includes also searching for quantitative trait loci using the method based on the mixed linear model.

The correlation networks and differential correlation networks were constructed to compare dependencies of metabolites under different conditions. Using WGCNA package in R, the Pearson correlation matrix was transformed into an adjacency matrix using a power function. Modules were detected by clustering the topological overlap matrix (TOM), which was used to create visualization of networks in Cytoscape. In order to find those metabolite associations that significantly differed between drought and control, differential correlation networks were created, using the test based on Fisher's Z transformation, with Bonferroni correction.

The algorithms can be adapted to any chromatographic data. Full description of the methods and the results are published in Piasecka et al. (2017).

Keywords: correlation networks, multifactorial experiments, REML, metabolomics

References

Piasecka A., Sawikowska A., Kuczyńska A., Ogródowicz P., Mikołajczak K., Krystkowiak K., Gudyś K., Guzy-Wróbelska J., Krajewski P. and Kachlicki P. 2017. *Drought-related secondary metabolites of barley (*Hordeum vulgare* L.) leaves and their metabolomic quantitative trait loci*. Plant Journal 89: 898–913.

Acknowledgements

This work was supported by the European Regional Development Fund through the Innovative Economy Program for Poland 2007-2013, project WND-POIG.01.03.01-00-101/08 POLAPGEN-BD „Biotechnological tools for breeding cereals with increased resistance to drought”.

**One and multi-variable characterization of spring barley
(*Hordeum vulgare L.*) cultivars grown in Smolice Plant Breeding
and tested in field-based breeding experiments in 2016**

TADEUSZ ŚMIAŁOWSKI^{1*}, ANNA CIEPLICKA², DARIUSZ R. MAŃKOWSKI¹

¹Laboratory of Seed Production and Plant Breeding Economics, Department of Seed
Science and Technology,
Plant Breeding and Acclimatization Institute
National Research Institute in Radzików

²Breeding Plants Smolice Ltd. Co. IHAR-PIB Group;
Departments Bąków

*e-mail: zhsmialo@cyf-kr.edu.pl

The aim of the study was to evaluate the variability of measured traits by the characteristics of the studied lines and varieties of spring barley. The research material was the lines of spring barley cultivated in the HR Smolice Breeding company, branch Bąków. These objects were tested in 2 series of team experiments in 6 localities (Bąków, Nagradowice, Polanowice, Radzików, Smolice and Strzelce). Field and laboratory results were the basis for statistical analyzes. The analysis of variance in cross-hierarchical model was used for the analysis of the diversity of the studied material in terms of yield, plant height, susceptibility and disease. The BWLUE estimators were estimated to evaluate the effects of the studied varieties and families. The self-organizing Kohonen neural network was used to develop the multi-character characteristics of the studied lines and varieties. The one-way analysis of variance for the yields obtained in the two series of experiments allowed us to conclude that in both series there were differences in the average yields obtained at each location. There was also a significant difference in average yields for the examined lines and varieties. Interaction between lines and varieties and locations was also statistically significant.

Keywords: comparative experiments, BWLUE estimators, plant breeding, spring barley, yielding

References

- Kohonen, T., 1982. *Self-organized formation of topologically correct feature maps*. Biological Cybernetics, Vol. 48, Issue: 1, 59-69.
- Mańkowski, D.R., 2013. Modele równań strukturalnych SEM w badaniach rolniczych, Monografie i Rozprawy Naukowe IHAR-PIB. IHAR-PIB, Radzików.

Prediction accuracy and consistency in cultivar ranking for factor-analytic linear mixed models for winter wheat multienvironmental trials

MARCIN STUDNICKI^{1*}, JAKUB PADEREWSKI¹, HANS PETER PIEPHO²,
ELŻBIETA WÓJCIK-GRONT¹

¹Department of Experimental Design and Bioinformatics
Warsaw University of Life Sciences, Poland

²Biostatistics Unit, Institute of Crop Science
University of Hohenheim, Germany

*marcin_studnicki@sggw.pl

In the majority of research on models for multienvironment trials, evaluation of the prediction accuracy of models with difference variance–covariance structures is focused on predicting the means for cultivar \times location ($C \times L$) combinations. In cultivar recommendation, however, it is often more important to evaluate prediction accuracy in modeling cultivar \times region ($C \times R$) combinations. The aim of this paper was to evaluate the prediction accuracy of two single-stage linear mixed models (LMMs) with different variance–covariance structures, emphasizing factor-analytic (FA) structures. One of the models was used to predict means for $C \times L$ combinations and the other one for $C \times R$ combinations. Additionally, we assessed implications of model choice for consistency in cultivar ranking. The data used for the analysis performed in this study were obtained from 42 locations and 47 winter wheat (*Triticum aestivum* L.) cultivars during three growing seasons within the Polish Post-Registration Variety Testing System. The data were assigned to six agroecological regions. For evaluating the prediction accuracy of LMMs, we used cross validation based on a modified equation for the mean squared error of prediction. Yield rankings modeled by different variance–covariance structures were compared by Spearman’s rank correlation. For each model with a different variance–covariance structure, we calculated the correlation coefficients between estimated and observed data. The model with the highest predictability for means of the $C \times L$ classification was the FA(2) variance–covariance structure. In the case of $C \times R$ means, the compound symmetry structure fared favorably, and using more complex variance–covariance structures (including heterogeneous covariances) did not increase prediction accuracy.

Keywords: agro-ecological regions; cross validation; cultivar ranking; factor-analytic model; multi-environment trial; spatial model; variance-covariance structure; winter wheat

Acknowledgements

For the Polish Research Centre for Cultivar Testing (COBORU) for providing the winter wheat MET data. H.P. Piepho was supported by DFG grant PI 377/17-1.

Likelihood function applied to detect dependency in 2×2 contingency tables

PIOTR SULEWSKI

Institute of Mathematics
Pomeranian Academy in Slupsk, Poland
piotr.sulewski@apsl.edu.pl

The paper focuses on 2×2 contingency tables (CTs). Scenarios of generation of CTs with the probability flow parameter are proposed and measures of untruthfulness of H_0 are defined. Statistic tests including the power divergence tests and the $|\chi|$ test under mentioned scenarios were defined and compared with maximum likelihood test in terms of their power. This paper is a simply attempt of replacing a nonparametric statistical inference method by the parametric one. Maximum likelihood method is applied to estimate the probability flow parameter. Instructions to generate CTs by means of the bar method are described. Numerical examples are presented and general conclusions are given.

Keywords: statistical inference, likelihood function, contingency tables, parametric test, probability flow parameter

References

- Blitzstein J., & Diaconis P. 2011. *A sequential importance sampling algorithm for generating random graphs with prescribed degrees*. Internet mathematics, 6(4):489-522.
- Bock H. H. 2003. *Two-way clustering for contingency tables: maximizing a dependence measure*. Between data science and applied data analysis, 143-154.
- Cressie N., Read T. R. 1984. *Multinomial goodness-of-fit tests*. Journal of the Royal Statistical Society. Series B (Methodological), 440-464.
- DeSalvo S., Zhao J. Y. 2015. *Random Sampling of Contingency Tables via Probabilistic Divide-and-Conquer*. arXiv preprint arXiv:1507.00070.
- Fishman G. S. 2012. *Counting contingency tables via multistage Markov chain Monte Carlo*. Journal of Computational and Graphical Statistics, 21(3):713-738.
- Sulewski P. 2009. *Two-by-two contingency table as a goodness-of-fit test*. Computational Methods in Science and Technology, 15(2):203-211.
- Sulewski P., Motyka R. 2015. *Power analysis of independence testing for contingency tables*. Scientific Journal of Polish Naval Academy, 56.

Implementation of testing the parallelism for four-parameter logistic model in R

ALICJA SZABELSKA BERĘSEWICZ, JOANNA ZYPRYCH-WALCZAK,
IDZI SIATKOWSKI*

Department of Mathematical and Statistical Methods,
Poznań University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland

*idzi@up.poznan.pl

It is often important for scientists to compare the potencies of two preparations, typically to determine potency of a test preparation relative to a standard preparation. It involves the testing of similarity between a pair of dose-response curves of reference standard and test sample. Typically, the four-parameter nonlinear logistic curve, are commonly used to model dose-response data. Traditionally, the answer about parallelism of two curves based on testing the hypothesis of equal parameters between the two dose-response curves. A typical approach is to perform F test of the null hypothesis that the relevant parameters are equal for the two dose-response curves (Gottschalk and Dunn, 2005). In this paper, we present an alternative method for testing parallelism in the four-parameter logistic response curve, based on an intersection union test presented in Jonkman and Sidik (2009).

Keywords: dose-response curve, four parameter logistic, nonlinear model, parallelism

References

- Gottschalk, P. G., Dunn, J. R. 2005. *Measuring parallelism, linearity, and relative potency in bioassay and immunoassay data*. J. Biopharm. Statist. 15, 437–463.
- Jonkman J. N, Sidik, K. 2009. *Equivalence testing for parallelism in the four-parameter logistic model*. J Biopharma Stat. 19, 818-37.

A double sigmoid curve for approximation of the biogenic amine production curve

MARTIN TLÁSKAL^{1*}, JAROSLAV MICHÁLEK²

¹Central Institute for Supervising and Testing in Agriculture, National Plant Variety Office,
Brno, Czech Republic

²Department of Econometrics, Faculty of Military Leadership, University of Defence in Brno,
Czech Republic

*martin.tlaskal@ukzuz.cz

Biogenic amines are compounds with a possible toxic effect. They can be found in foodstuffs and beverages at different concentrations. Intake of these compounds in higher amount can lead to some undesirable health difficulties of the consumer. In order to approximate the time course of biogenic amine production, simple sigmoid functions can be used. However, they give unsatisfactory fit in the case of biphasic production.

The aim of this paper is to introduce a model that is able to describe double sigmoid pattern observed in data. In order to obtain such a model, higher derivatives of the Gompertz curve were exploited. As far as we know, this approach has not been used for obtaining double sigmoid functions. Some properties of the model as well as advantages and drawbacks of its practical use are given. Possible biological explanations are also suggested. Usefulness of the model is demonstrated on chosen data sets describing tyramine production.

Keywords: biogenic amines, biphasic growth, double sigmoid, Gompertz curve, modelling, tyramine

References

- Hau B., Amorim L., Bergamin Filho A. 1993. *Mathematical functions to describe disease progress curves of double sigmoid pattern*. *Phytopathology* 83: 928–932.
- Lipovetsky S. 2010. *Double logistic curve in regression modeling*. *Journal of Applied Statistics* 37: 1785–1793.
- Meyer P. 1994. *Bi-logistic growth*. *Technological Forecasting and Social Change* 47: 89–102.
- Silla Santos M. 1996. Biogenic amines: Their importance in foods. *International Journal of Food Microbiology* 29: 213–231.

Estimation methods of general and specific combining ability effects for non-orthogonal data

KRZYSZTOF UKALSKI, JOANNA UKALSKA

Biometry Division, Department of Econometrics and Statistics, Faculty of Applied
Informatics and Mathematics, Warsaw University of Life Sciences

The purpose of the study was to compare four estimation methods of general (GCA) and specific combining ability (SCA) effects for incomplete North Carolina II design. We compared results obtained by: Ubysz-Borucka et al. (1985, the method of generalized inverse matrix), Trętowski and Wójcik (1988, the iterative method), Laudański (1996, the method of projective operators) with the restricted maximum like-lihood (REML) method in SAS program. The example chosen is an incomplete diallel cross of six tulip cultivars (Garretens and Keuls, 1978). The data used are the bulb weights of tulip seedlings in the first year after sowing.

Keywords: general combining ability, specific combining ability, non-orthogonal data, generalized inverse, projective operators, restricted maximum like-lihood

References

- Garretens F., Keuls M. 1978. *A general method for the analysis of genetic variation in complete and incomplete diallels and North Carolina II design. Part II. Procedures and general formulas for the fixed model.* Euphytica 27: 49-68.
- Laudański Z. 1996. *Zastosowanie operatorów rzutowania w analizowaniu danych nieortogonalnych.* Wydawnictwo SGGW.
- Trętowski J., Wójcik A.R. 1988. *Metodyka doświadczeń rolniczych.* WSPR Siedlce.
- Ubysz-Borucka L., Mądry W., Muszyński S. 1985. *Podstawy statystyczne genetyki cech ilościowych w hodowli roślin.* Wydawnictwo SGGW-AR.

**Response of gas exchange to leaf piercing explained by piecewise
linear regression for two developmental forms of rape plant
(*Brassica napus* L. ssp. *oleifera* Metzg)**

ANNA WENDA-PIESIK^{1*}, MACIEJ KAZEK¹, WŁODEK KRZESIŃSKI²

¹Department of Plant Growth Principles and Experimental Methodology, UTP University of Science and Technology in Bydgoszcz, Poland

² Department of Vegetable Crops, University of Life Sciences in Poznań, Poland

*apiesik@utp.edu.pl

Oilseed rape (*Brassica napus* L. ssp. *oleifera* Metzg) was the subject of the study in two forms: winter cv. ‘Muller’ (at the rosette stage – the first internode BBCH 30 – 31) and spring cv. ‘Feliks’ (at the yellow bud stage BBCH 59). The main gas-exchange parameters, net photosynthetic rate (P_N), transpiration rate (E), stomatal conductance (g_s), and intercellular CO₂ concentration (C_i) were measured on leaves prior to the piercing and immediately after the short-term piercing. The analysis of piecewise linear regression with the breakpoint estimation was calculated to explain the time-relation changes of the parameters g_s , C_i , P_N , and E . The time intervals were estimated by nonlinear methods according to Quasi-Newton and the lost function based on the least squares were proceeded (Haelterman et al., 2009). The relationships between the parameters were computed using the simple coefficient of correlation (r by Pearson).

The effect of mechanical wounding revealed different progress of the gas exchange process for the two forms. Piecewise linear regression with the breakpoint estimation showed that the plants at the same age but at a different vegetal stage, manage mechanical leaf-piercing differently. The differences concerned the stomatal conductance and transpiration changes since for rosette leaves the process consisted of five intervals with a uniform direction, while for stem leaves - of five intervals with a fluctuating direction. These parameters got stabilized within a similar time (220 mins) for both forms. The process of net photosynthetic rate was altered by the plant stages. ‘Muller’ plants at the rosette stage demonstrated dependence of P_N on time in *log-linear* progression: $y (P_N) = 8.01 + 2.73 \log_{10}(x t_2)$; $7 < t_2 < 220$; $R^2 = 0.96$. For stem leaves of ‘Feliks’ plants the process of transpiration, in terms of directions, was convergent with the process of photosynthesis. Those two processes were synchronized from 1st to 114th min of the test ($r = 0.85$; $p < 0.001$) in plants at the rosette stage and from 26th to 148th min in stem leaves ($r = 0.95$; $p < 0.001$).

Keywords: gas exchange parameters, wounding of leaf, piecewise linear regression

References

Haelterman R, Degroote J, van Heule D, Vierendeels J. 2009. *The quasi-Newton Least Squares method: a new and fast secant method analyzed for linear systems*. SIAM Journal on Numerical Analysis 47: 2347–2368.

Adopting Hellwig`s method for selecting concomitant variables at a certain growth curve model

MIROSLAWA WESOŁOWSKA-JANCZAREK, MONIKA RÓŻAŃSKA-BOCZULA*

Department of Applied Mathematics and Computer Science

University of Life Sciences in Lublin

Głęboka 28, 20-612 Lublin

*monika.boczula@up.lublin.pl

The paper presents the application of the Hellwig`s method for selecting concomitant variables at a growth curve model with concomitant variables the values of which change over time and are the same for all experimental units. The authors show simple adaptation of a part of the growth curves model to the multiple regression model for which the Hellwig`s method applies. Theoretical considerations are used for choosing significant concomitant variables for raspberries fructification.

Keywords: growth curve method, concomitant variables, Hellwig`s method

References

- Fujikoshi Y., Rao C.R. 1991. *Selection of covariables in the growth curve model*. *Biometrika* 78, 4: 779-785.
- Fus L., Wesołowska-Janczarek M. 1998. *Comparison of two growth curve models with time moving concomitant variables*. *Colloquium Biometricum* 28: 108-115.
- Hellwig Z. 1987. *Elementy rachunku prawdopodobieństwa i statystyki matematycznej*. PWN Warszawa.
- Wang S.-G., Liski E., Nummi T. 1999. *Two-way selection of covariables in multivariate growth curve models*. *Linear Algebra and its Applications* 289: 333-342.
- Wesołowska-Janczarek M., Fus L. 1996. *Parameters estimation in the growth curves model with time-changing concomitant variables*. In Polish. *Colloquium Biometricum* 26: 263-277.
- Wesołowska-Janczarek M., Fus L., Osypiuk Z. 1997. *Zastosowanie metody krzywych wzrostu ze zmiennymi towarzyszącymi do badania owocowania malin*. In Polish. *Colloquium Biometricum* 27: 269 – 281.

CI thermometer:

Visualizing confidence intervals in correlation analysis

AGNIESZKA WNUK^{1*}, KONRAD DEBSKI², MARCIN KOZAK³

¹Department of Experimental Design and Bioinformatics
Warsaw University of Life Sciences, Poland

²Laboratory of Bioinformatics, Neurobiology Center, Nencki Institute of Experimental
Biology, Polish Academy of Sciences, Poland

³Department of Quantitative and Qualitative Methods,
University of Information Technology and Management in Rzeszów, Poland

*kanea.oo@gmail.com

Used by so many researchers in so many scientific disciplines, correlation is everywhere. However, what we usually see is only estimates of the correlation coefficient. The numbers themselves far-too-often do not offer sufficient description of the studied phenomena, which is why many researchers feel like adding more information about the correlation, such as verification of the hypothesis that the correlation coefficient is null. This is presented with asterisks (e.g., “*” represents correlation coefficient significant at $P \leq 0.05$), actual p -values, or (far less often) confidence intervals. These methods are easy if one reports a single correlation coefficient, but for many coefficients, often reported in a correlation matrix, the methods become difficult. Correlation tables with p -values and confidence intervals are difficult to read, so one is stuck asterisks, an overly simplistic method. We offer a new visualization technique called “CI thermometer”, which will help scientists present correlation matrices accompanied by additional information on confidence intervals.

Keywords: correlation matrix, corrplot, significance, visualization

Modeling the yield-scaled Global Warming Potential for winter wheat production in Poland

ELŻBIETA WÓJCIK-GRONT^{*}, MARCIN STUDNICKI

Department of Experimental Design and Bioinformatics
Warsaw University of Life Sciences – SGGW, Poland

*ewojcik.gront@gmail.com

This paper presents the results of modeling the yield-scaled Global Warming Potential (GWP) for winter wheat production in Poland using three models: linear mixed models (LMMs) (Patterson and Thompson 1971, Searle et al., 1992), classification and regression tree (CART) and random forest (RF) (Breiman et. al., 1984). The main objective of this study was to determine major drivers of variation in yield-scaled GWP of winter wheat in East-central Europe focusing on climatic, environmental and genetic variables. The yield-scaled GWP is the Global Warming Potential calculated per grain unit expressed in kg CO₂ equivalent kg⁻¹ yield of winter wheat (Wójcik-Gront and Bloch-Michalik, 2016). The GWP of winter wheat production is estimated in Life Cycle Assessment (LCA) (Consoli et al. 1993). In this study both models: the regression trees based analysis and LMM revealed that location is the most influential predictors of the yield-scaled GWP variability in winter wheat production.

Keywords: GHG emission; variability; multivariate analysis, regression trees, linear mixed models

References

- Breiman, L. , J. H. Friedman , R. A. Olshen , and C. G. Stone. 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, California, USA.
- Consoli, F., Allen, D, Boustead, I., Fava, J., Franklin, W., Jensen, A.A., de Oude, N., Parrish, R., Perriman, R., Postlethwaite, D., Quay, B., Séguin, J., Vignon, B., 1993. (Eds.) *Guidelines for Life-Cycle Assessment: A 'Code of Practice'*. Society of Environmental Toxicology and Chemistry (SETAC), Brussels.
- Patterson, H.D., and R. Thompson. 1971. *Recovery of interblock information when block sizes are unequal*. *Biometrika* 58:545–554.
- Searle, S.R., G. Casella, and C.E. McCulloch, editors. 1992. *Variance components*. John Wiley & Sons, New York.
- Wójcik-Gront, E., Bloch-Michalik, M. 2016. *Assessment of greenhouse gas emission from life cycle of basic cereals production in Poland*. *Zemdirbyste-Agriculture* 103 (3): 259–266.

Using a mixed-level fractional factorial design to evaluate the effect of the intensity of agronomic practices on the yield of different white mustard (*Sinapis alba* L.) morphotypes

DARIUSZ ZAŁUSKI^{1*}, KRZYSZTOF JANKOWSKI²

¹Department of Plant Breeding and Seed Production
University of Warmia and Mazury in Olsztyn, Poland

²Department of Agrotechnology, Agricultural Production Management and Agribusiness
University of Warmia and Mazury in Olsztyn, Poland

*dariusz.zaluski@uwm.edu.pl

Significant progress in plant breeding and molecular genetics contributed to the development of white mustard (*Sinapis alba* L.) genotypes/cultivars whose biomass (mainly seeds) can be used for a variety of purposes. Those advancements are also used to modify production technologies. In systems with many production factors, agronomic measures exert direct and indirect (interaction) effects on crops and yield. Multiple production factors can be analyzed simultaneously with the use of fractional factorial s^{k-p} designs. A field experiment a mixed two-level and three-level fractional factorial resolution V design with $k=5$ factors was performed at the Agricultural Experiment Station in Bałcyny (north-eastern Poland), owned by the University of Warmia and Mazury in Olsztyn. The study investigated the responses of two cultivars (traditional and canola) of white mustard (*Sinapis alba* L.) to the main seeds yield-forming factors (seeding date, nitrogen fertilization, sulfur fertilization, boron fertilization).

The results can optimize decision-making and constitute a basis for establishing effective levels of key operations in the production of new morphotypes and genotypes of agricultural crops.

Keywords: experimental design, field experiment, resolution V

Comparison of parameters of wheat flour dough obtained from Extensograf Farinograph and Glutapeak

BOGNA ZAWIEJA^{1*}, AGNIESZKA MAKOWSKA², MATEUSZ GUTSCHE³

¹Department of Mathematical and Statistical Methods,
Poznan University of Life Sciences, Poland

²Institute of Food Technology of Plant Origin,
Poznan University of Life Sciences, Poland

³GoodMills Polska Spółka z o.o., Poznań.

*bogna13@up.poznan.pl

Each batch of grain supplied to the mill is subjected to quality control of the flour obtained from it. For this the dough parameters are measured. Among them water absorption, dough stability and energy are evaluated. The first two parameters are usually evaluated by Farinograph, while the third one by Extensograph, but these methods are time-consuming. In recent years, the new instrument, GlutaPeak, has appeared on the market. The measurement of flour parameters via this instrument takes about 10 minutes. The following parameters are obtained from this instrument: peak maximum time, torque maximum, torque before maximum, torque after maximum and specific areas. Millers would like to know the answer to whether the parameters obtained from the new instrument are correlated with parameters obtained from conventional instrumentation. In addition, they need an algorithm to recalculate the data derived from the new instrument to parameters that they are already acquainted with.

The first the linear correlation coefficients (CV) between the analyzed parameters were analyzed. Next sequential (backward, forward and stepwise) multiple quadratic regression analysis, logistic regression analysis and PLS regression analysis was applied. The models best suited for data were selected using the AIC criterion. The CV between the water absorption of wholegrain flour with salt and a few parameters obtained from GlutaPeak were higher than 75%. For stability, the highest value of CV was 40% and 44% for energy. In the case of water absorption, the best fit was the square regression model and for the dough stability and energy the best matched models were logistics regression models.

Based on the conducted analyzes, it can be stated that the water absorption can be assessed, without great error, on the basis of the results of the measurements performed with GlutaPeak. The results obtained from logistic models for energy and stability, even though not exceptionally good, can be used with some caution ("flattening" results could be expected: the flour is very rarely classified as extreme classes, which could be due to the fact that the number of data in these classes was small).

Keywords: dough stability, energy mill, logistic regression, PLS, water absorption,