



51 Międzynarodowe Colloquium Biometryczne  
dedykowane

Panu Profesorowi Mirosławowi Krzyśko  
z okazji 80–lecia urodzin

Szamotuły, 11–14 września 2022

POLSKIE TOWARZYSTWO BIOMETRYCZNE

# Construction method of an incomplete split-plot $\times$ split-block design using some resolvable block designs

KATARZYNA AMBROŻY-DERĘGOWSKA<sup>1\*</sup>

<sup>1</sup>Department of Mathematical and Statistical Methods  
Poznań University of Life Sciences in Poznań, Polska

\*katarzyna.ambrozy@up.poznan.pl

The construction of incomplete split-plot  $\times$  split-block (SPSB) design for three factor experiments is considered. A semi-Kronecker (Khatri-Rao) product of matrices is used in the construction procedure. With this method some generated design is obtained when column treatments and subplot treatments are allocated in balanced square lattice designs while row treatments are in some resolvable incomplete block design. There are considered experiments with an orthogonal block structure. Stratum efficiency factors are given when the split-plot  $\times$  split-block design is generally balanced.

Keywords: General balance property, Efficiency factors, Incomplete split-plot  $\times$  split-block designs, Khatri–Rao product of matrices, Resolvable design

## References

- Ambroży K., Mejza I. (2003): Construction of split-plot  $\times$  split-block designs using balanced square lattice designs. *Colloquium Biometryczne*, 33a, 7–19.
- Caliński, T., Kageyama, S. (2000). *Block Designs. A Randomization Approach*, Volume I. *Analysis*. Lecture Notes in Statistics 150, Springer-Verlag, New York.
- Caliński, T., Kageyama, S. (2003). *Block Designs. A Randomization Approach*, Volume II. *Design*. Lecture Notes in Statistics 170, Springer-Verlag, New York.
- Houtman A.M., Speed T.P. (1983): Balance in designed experiments with orthogonal block structure. *Ann. Statist.* 11, 1069-1085.
- Khatri C.G., Rao C. R. (1968): Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhya Ser. A* 30,167-180.
- Ozawa K., Mejza I., Mejza S., Kuriki S. (2018): Two-factor experiments with split units constructed by cyclic designs and square lattice designs. *Statistical Journal*, Volume 16, Number 3, 279-294.

# Comparison of interpolation methods for mapping soil properties

URSZULA BRONOWICKA-MIELNICZUK<sup>\*1</sup>, JACEK MIELNICZUK<sup>1</sup>

<sup>1</sup>Department of Applied Mathematics and Computer Science  
University of Life Sciences in Lublin, Poland  
<sup>\*</sup>urszula.bronowicka@up.lublin.pl

The choice of the optimal interpolation technique for estimating soil properties is an important issue from both an agricultural, ecological and economic perspective. The purpose of this study was to evaluate deterministic and stochastic interpolation methods to selected soil properties collected from the LUCAS Soil 2015 dataset which is freely available by the European Commission through the European Soil Data Centre managed by the Joint Research Centre (JRC, <http://esdac.jrc.ec.europa.eu/>). In the study we compared eight interpolation methods such as triangulation with linear interpolation (TLI), inverse distance weighting (IDW), modified Shepard's method (MS), nearest neighbour interpolation (NeN), natural neighbour interpolation (NaN), local polynomials interpolation (LPI), radial basis function (RBF) and ordinary kriging (OK). We shall quantify and compare the quality of interpolation models by means of classical measures and some new indices. Spatial interpolations were performed in Surfer ver. 23 and ArcGIS 10.6.1, other calculations were conducted using R software (ver. 4.1.0).

**Keywords:** cross-validation, interpolation methods, soil properties

## References

- Cressie, N. 1993. *Statistics for spatial data*. John Wiley & Sons, Inc.
- Isaaks E.H., Srivastava R.M. 1989. *An introduction to applied geostatistics*. Oxford University Press, Oxford.
- Jones, A, Fernandez-Ugalde, O., Scarpa, S. *LUCAS 2015 Topsoil Survey. Presentation of dataset and results*, EUR 30332 EN, Publications Office of the European Union: Luxembourg. 2020
- Wackernagel H. 1998. *Multivariate geostatistics. An Introduction with applications*. 2nd completely revised edition. Springer.
- Webster R., Oliver M. 2001. *Geostatistics for environmental scientists*. John Wiley & Sons, Ltd, Chichester.
- Willmott C.J. 1981. *On the validation of models*. *Physical Geography* 2: 184-194.

## **Metody uzupełniania brakujących obserwacji w badaniach jakości powietrza**

DOROTA DOMAGAŁA<sup>1\*</sup>, MAŁGORZATA SZCZEPANIK<sup>1</sup>, AGNIESZKA KUBIK-KOMAR<sup>1</sup>

<sup>1</sup>Katedra Zastosowań Matematyki i Informatyki  
Uniwersytet Przyrodniczy w Lublinie  
<sup>\*</sup>[dorota.domagala@up.lublin.pl](mailto:dorota.domagala@up.lublin.pl)

Podczas analizy wyników eksperymentu możemy mieć do czynienia z występowaniem brakujących obserwacji, co może doprowadzić do wyciągnięcia błędnych wniosków z przeprowadzonego badania. Istotny jest zatem sposób postępowania w takiej sytuacji. W zależności od charakteru danych oraz typu brakujących obserwacji możliwe jest zastosowanie różnych metod uzupełniania braków: prostych lub złożonych. W pracy zastosowano i porównano wybrane techniki imputacji brakujących danych w analizie parametrów jakości powietrza (poziomu pyłów zawieszonych PM<sub>2.5</sub> oraz PM<sub>10</sub>, a także wybranych zanieczyszczeń gazowych).

Słowa kluczowe: brakujące obserwacje, imputacja brakujących obserwacji, jakość powietrza.

## Catalogue of selected experimental plans

MAŁGORZATA GRACZYK<sup>\*1</sup>, BRONISŁAW CERANKA<sup>1</sup>

<sup>1</sup>Department of Mathematical and Statistical Methods  
Poznan University of Life Sciences  
<sup>\*</sup>[malgorzata.graczyk@up.poznan.pl](mailto:malgorzata.graczyk@up.poznan.pl)

There are many works in the literature on the construction of experimental plans based on weighing designs. Hence, it was necessary to designate a catalog of experimental designs with specific properties. In this work, we investigate the properties of experimental plans constructed by used of the matrices of spring weighing designs. Additionally, an even number of experimental objects was assumed. An overview of the construction methods of these designs and classes of design matrices with selected properties are given. The obtained results allowed to create a catalogue of experimental designs constructed on the basis of spring weighing designs.

Keywords: plan of experiment, spring balance weighing design

### References

- Ceranka B., Graczyk M. (2014). Regular D-optimal spring balance weighing designs: construction. *Acta Universitatis Lodzianensis, Folia Oeconomica* 302, 111-125.
- Ceranka B., Graczyk M. (2017). Highly D-efficient weighing design and its construction. *Acta Universitatis Lodzianensis, Folia Oeconomica* 331, 143–151.
- Ceranka B., Graczyk M. (2018). Highly D-efficient weighing designs for an even number of objects. *Revstat Statistical Journal* 16(4), 475–486.
- Katulska K., Przybył K. (2007). On certain D-optimal spring balance weighing designs. *Journal of Statistical Theory and Practice* 1, 393–404.

# Comics as a means of teaching and propagating statistics

JOLANTA GRALA-MICHALAK<sup>1\*</sup>

<sup>1</sup>Faculty of Mathematics and Computer Science  
Adam Mickiewicz University in Poznań, Poland  
\*grala@wmi.amu.edu.pl

Comics are an element of mass culture, a means of expression combining text and drawings. In recent years, the creation of comic book stories has been included in the curriculum of primary schools in Poland, as a task to be performed by students on the subject: language Polish, computer science, education for safety, and others. Elements of descriptive statistics were included in the core curriculum of the subject of mathematics in grades 7 and 8. Comic book "Kajko and Kokosz. School of Flying" went to the canon of compulsory reading in grade 6. The following question arises: can statistical issues and problems be (and should) be presented in the language of comics? The affirmative answer leads to further questions: what limitations and difficulties does this form of creativity impose on the author? What can be the reception of this form of knowledge transfer and to whom is it to be addressed? During the lecture there will be partial answers to the problems posed. First, we will focus on presenting examples of book publications of a comic book character or elements, intended for Polish readers. Then the features of a good statistical comic will be given. The speech will end with an encouragement to create such works as complementary reading for students.

Keywords: academic textbooks, comic book stories, new trends in teaching statistics

## References

- Biecek, P., Baranowska E., Sobczyk, P. 2019. *Wykresy unplugged*. Fundacja Naukowa Smarter Poland, Warszawa.
- Gonick, L., Smith, W. 2011. *Przewodnik MOCNO ilustrowany*. Wolters Kluwer.
- Gonick, L., Smith, W. 1993. *The Cartoon Guide to Statistics*. HarperPerennial.
- Klein, G., Dabney, A. 2018. *Komiksowe wprowadzenie do statystyki*. Wydawnictwo Naukowe PWN S.A., Warszawa.
- Klein, G., Dabney, A. 2013. *The Cartoon Introduction to Statistics*. Hill and Weng Inc., U.S.
- Magnello, E., Van Loon, B. 2009. *Introducing Statistics*. Icon Books Ltd, London.
- Pawlak, R. 2019. *On teacher education and on key problems of teaching mathematics*. *Didactica Mathematicae*, North America, 41:123-134.
- Santos, V., Pereira, A., Neto, T., M. Pinheiro, M. 2022. *The Math Anxiety Project at Higher Education: A Case Study from Portugal*. *Didactica Mathematicae*, North America, 43: 89-99.

# Empirical power of a modified normalizing transformation statistic for testing multivariate normality based on kurtosis

ZOFIA HANUSZ<sup>1\*</sup>, ERI KURITA<sup>2</sup> AND TAKASHI SEO<sup>2</sup>

<sup>1</sup>DEPARTMENT OF APPLIED MATHEMATICS AND COMPUTER SCIENCE,  
UNIVERSITY OF LIFE SCIENCES IN LUBLIN, POLAND

<sup>2</sup>DEPARTMENT OF APPLIED MATHEMATICS, TOKYO UNIVERSITY OF SCIENCE,  
JAPAN

\*zofia.hanusz@up.lublin.pl

The paper deals with testing multivariate normality of data sets. In Enomoto et al. (2020) the normalizing statistic for Mardia's sample measure of multivariate kurtosis has been proposed. In the paper of Kurita et al. (2022), some modification of this normalizing transformation test statistic has been presented. This test statistic improves the normal approximation by using the exact expectation and variance of Mardia's multivariate sample kurtosis. The accuracy of the test statistic, namely, the expectation, variance, skewness, kurtosis and sample significance level were checked in simulation study. In the present study, power of the proposed test statistic in Kurita et al. (2022) is calculated and is compared with power of the normalizing transformation statistic in Enomoto et al. (2020). Moreover, the power results have been also compared to a more accurate Mardia's test statistic based on kurtosis (Mardia, 1994), and the Henze-Zirkler test (Henze and Zirkler, 1990), recommended as a formal test for multivariate normality. Empirical power was calculated via Monte-Carlo simulations for chosen alternative distribution characterizing by different deviations from multivariate normality. All calculations were done in Mathematica 11, for 10,000 random data sets generated for combinations of number of sizes  $n = 20, 30, 40, 50, 100$  and dimensions  $p = 2, 3, 4, 5, 7, 10$  at significance level 0.05.

## REFERENCES

- [1] Enomoto R., Hanusz Z., Hara A., Seo T. (2020). Multivariate normality test using normalizing transformation for Mardia's multivariate kurtosis. *Communication in Statistics - Simulation and Computation*, 49, 684-698.
- [2] Henze N., Zirkler B. (1990). A class of invariant consistent tests for multivariate normality. *Communication in Statistics - Theory and Methods*, 19, 3595-3617.
- [3] Mardia K.V. (1974). Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. *Sankhya B*, 36, 115-128.
- [4] Kurita E., Seo T., Hanusz Z. (2022). A modified normalizing transformation statistic for multivariate normality test based on kurtosis (in printing process)

## Spatio-temporal principal component analysis based on Geary's contiguity ratio

MIROSLAW KRZYŚKO<sup>1\*</sup>, PETER NIJKAMP<sup>2</sup>, WALDEMAR RATAJCZAK<sup>3</sup>, WALDEMAR  
WOŁYŃSKI<sup>4</sup>, ANDRZEJ WOJTYŁA<sup>1</sup>, BEATA WENERSKA<sup>1</sup>

<sup>1</sup>Calisia University-Kalisz, Poland

<sup>2</sup>Open University, Heerlen, The Netherlands

<sup>3</sup>Faculty of Socio-Economic Geography and Spatial Management,  
Adam Mickiewicz University, Poznań, Poland

<sup>4</sup>Faculty of Mathematics and Computer Science,  
Adam Mickiewicz University, Poznań, Poland

\*mkrzysko@amu.edu.pl

Principal Component Analysis (PCA) is an established research method in the quantitative social sciences. The present study aims to construct and test a new methodology for a proper treatment of spatio-temporal dependencies among data in a PCA, based on recent advances in multifunctional PCA. In our approach, the product of the functional principal component variance and Geary's contiguity ratio  $C$  is optimized. The analytical significance of spatio-temporal principal components is confirmed by an application to a spatial multitemporal data set concerning the development of Polish regions in the years 2002-2018, as characterized by 12 socioeconomic features.

Keywords: Geary's contiguity ratio  $C$ , functional data, spatio-temporal data, spatio-temporal principal components

### References

- Demas, U., Harris, P., Brundson, P., Fotheringham, A.S., McLoone, S. 2013. *Principal component analysis on spatial data: An overview*. Annals of the Association of American Geographers 103 (1): 106-128.
- Geary, R.C. 1954. *The contiguity ratio and statistical mapping*. The Incorporated Statistician 5: 115-145.
- Hotelling, H. 1933. *Analysis of a complex of statistical variables into principal components*. Journal of Educational Psychology 24: 417-441.
- Krzyśko, M., Nijkamp, P., Ratajczak, W., Wołyński, W. 2022. *Multidimensional Economic Indicators and Multivariate Functional Principal Components Analysis*. Journal of Geographical Systems 24(1): 49-65.



# Application of distance standard deviation in functional data analysis

MIROSLAW KRZYŚKO<sup>1</sup>, ŁUKASZ SMAGA<sup>2\*</sup>

<sup>1</sup>Interfaculty Institute of Mathematics and Statistics  
Calisia University-Kalisz, Kalisz, Poland

<sup>2</sup>Faculty of Mathematics and Computer Science  
Adam Mickiewicz University, Poznań, Poland

[\\*ls@amu.edu.pl](mailto:*ls@amu.edu.pl)

This paper concerns the measurement and testing of equality of variability of functional data. We apply the distance standard deviation constructed based on distance correlation, which was recently introduced as a measure of spread. For functional data, the distance standard deviation seems to measure different kinds of variability, not only scale differences. Moreover, the distance standard deviation is just one real number, and for this reason, it is of more practical value than the covariance function, which is a more difficult object to interpret. For testing equality of variability in two groups, we propose a permutation method based on centered observations, which controls the type I error level much better than the standard permutation method. We also consider the applicability of other correlations to measure the variability of functional data. The finite sample properties of two-sample tests are investigated in extensive simulation studies. We also illustrate their use in real data examples based on various data sets.

Keywords: distance correlation, functional data analysis, projection correlation, standard deviation, variability measure

## References

- Edelmann D., Richards D., Vogel D. 2020. *The distance standard deviation*. The Annals of Statistics 48: 3395-3416.
- Horváth L., Kokoszka P. 2012. Inference for functional data with applications. Springer.
- Ramsay J.O., Silverman B.W. 2005. Functional data analysis. Second Edition. Springer.
- Székely G.J., Rizzo M.L., Bakirov N.K. 2007. *Measuring and testing dependence by correlation of distances*. The Annals of Statistics 35: 2769-2794.

## Detekcja ziaren pyłku na obrazach mikroskopowych

ELŻBIETA KUBERA<sup>1\*</sup>, AGNIESZKA KUBIK-KOMAR<sup>1</sup>, PAWEŁ KURASIŃSKI<sup>1</sup>,  
 KRYSZYNA PIOTROWSKA-WERYSZKO<sup>2</sup>, MAGDALENA SKRZYPIEC<sup>3</sup>

<sup>1</sup>Katedra Zastosowań Matematyki i Informatyki  
 Uniwersytet Przyrodniczy w Lublinie, Polska

<sup>2</sup>Katedra Botaniki i Fizjologii Roślin  
 Uniwersytet Przyrodniczy w Lublinie, Polska

<sup>3</sup>Instytut Matematyki  
 Uniwersytet Marii Curie-Skłodowskiej, Lublin, Polska  
 \*elzbieta.kubera@up.lublin.pl

Analiza materiału pyłkowego pozyskanego z aparatu typu Hirsta jest zwykle przeprowadzana pod mikroskopem przez specjalistę palinologa. Jest to pracochłonne i czasochłonne zadanie. Celem tych badań jest automatyzacja procesu detekcji i zliczania ziaren pyłku różnych taksonów na podstawie cyfrowych obrazów mikroskopowych. Najpopularniejsze obecnie detektory obiektów - sieci YOLO, Faster R-CNN i RetinaNet były uczone detekcji ziaren pyłku z obrazów mikroskopowych materiału referencyjnego zawierającego trzy taksony typowe dla Europy Środkowej i Wschodniej. Najlepszy z zastosowanych detektorów umożliwia uzyskiwanie detekcji na poziomie mAP@.5:95 około 90%. Miara ta, stosowana do porównywania mechanizmów detekcji, wskazuje dokładność zarówno lokalizacji jak klasyfikacji obiektów.

Keywords: deep neural networks, detection, pollen monitoring

### References

- Holgate, S.T., Church, M.K., Lichtenstein, L.M. 2001 Allergy. 2nd. ed. Mosby International Ltd., London.
- Hirst, J.M. 1952. An Automatic Volumetric Spore Trap. *Ann. Appl. Biol.*, 39, 257–265, doi:10.1111/J.1744-7348.1952.TB00904.X
- LeCun, Y., Bengio, Y., Hinton, G. 2015. Deep Learning. *Nature*, 521, 436–444, doi:10.1038/nature14539
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. 2016. You only look once: unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*
- Ren S, He K, Girshick R, Sun J. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, pp. 1137-1149

## Statistical data analysis in plant experiments based on RNA-seq technology

GRAŻYNA NIEDZIELA\*, ALICJA SZABELSKA-BERĘSEWICZ,  
JOANNA ZYPRYCH-WALCZAK, MAŁGORZATA GRACZYK

Department of Mathematical and Statistical Methods  
Poznań University of Life Sciences, Poland  
\*grazyna.niedziela@up.poznan.pl

RNA-seq technology based on transcriptome sequencing is particularly used in differential expression studies. Many analyzes using RNA-seq technologies relate to medical or animal data. There are far fewer plant datasets in publicly available RNA-seq data repositories. Therefore, during the presentation, statistical analyzes will be presented for a number of plant data (Cantu et al. 2011; Pass et al. 2017; Frąckowiak et al. 2022; Wielkopolan et al. 2022). There are many statistical methods used for seeking genes that differentiate between two or more conditions, causing scientists to often ask which method to use? In an attempt to answer this question, the two most popular methods of differential analysis will be presented: edgeR (Robinson et al. 2010) and DESeq (Anders and Huber, 2010). The performed differential analysis with usage of these two methods on the presented datasets allowed for their comparison. The conclusions drawn from the comparative analysis will support the scientists' decisions regarding the choice of method.

Keywords: edgeR, DESeq, RNA-seq, statistical methods for differential analysis

### References

- Anders, S. and Huber, W. 2010. *Differential expression analysis for sequence count data*. Genome biology, 11(10):R106.
- Cantu, D., Pearce, S. P., Distelfeld, A., Christiansen, M. W., Uauy, C., Akhunov, E., Fahima, T., and Dubcovsky, J. 2011. *Effect of the down-regulation of the high grain protein content (gpc) genes on the wheat transcriptome during monocarpic senescence*. BMC genomics, 12(1):1–17.
- Frąckowiak, P., Wrzesinska, B., Wiczorek, P., Sanchez-Bel, P., Kunz, L., Dittmann, A., and Obrepalska-Stęplowska, A. 2022. *Deciphering of bth-induced response of tomato (solanum lycopersicum l.) and its effect on plant virus infection through the multi-omics approach*. bioRxiv.
- Pass, D. A., Sornay, E., Marchbank, A., Crawford, M. R., Paszkiewicz, K., Kent, N. A., and Murray, J. A. 2017. *Genome-wide chromatin mapping with size resolution reveals a dynamic sub-nucleosomal landscape in arabidopsis*. PLoS Genetics, 13(9):e1006988.
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. 2010. *edgeR: a bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics, 26(1):139–140.
- Wielkopolan, B., Frąckowiak, P., Wiczorek, P., and Obrepalska-Stęplowska, A. (2022). *The impact of oulema melanopus—associated bacteria on the wheat defense response to the feeding of their insect hosts*. Cells, 11(15):2342.

## Active protection of endangered species of peat bog flora (*Drosera intermedia*, *D. anglica*) in the Łęczna-Włodawa Lake District

JOANNA SENDER<sup>1\*</sup>, MONIKA RÓŻAŃSKA-BOCZULA<sup>2</sup>, DANUTA URBAN<sup>3</sup>

<sup>1</sup> Department of Hydrobiology and Protection of Ecosystems, University of Life Sciences in Lublin, 20-262, Lublin, Poland

<sup>2</sup> Department of Applied Mathematics and Computer Science  
University of Life Sciences in Lublin, Poland

<sup>3</sup> Institute of Soil Science and Environmental Engineering and Management, Department of Biological Bases of Forestry, University of Life Sciences in Lublin, 20-069, Lublin, Poland\  
\*joanna.sender@up.lublin.pl

Protecting endangered plant species is crucial to maintaining biodiversity. Currently, there is a rapid decline in the populations of many moisture-loving plant species throughout Poland. These include the entire genus *Drosera*, so the paper attempts to determine the main reasons for the decline of *Drosera intermedia* and *D. anglica* in the Leczna-Wlodawa Lake District. For this purpose, the habitat conditions of the sites of their current and historically abundant occurrence were compared. The hierarchical cluster analysis method, together with the Calinski-Harabasz index, was used to evaluate differences between sites in terms of changes in sundew frequency. The similarity between sites was represented using a dendrogram with a design based on the Bray-Curtis dissimilarity index. Principal component analysis was used to characterise spatial variation in physical and chemical parameters. For *Drosera anglica*, a clear preference was observed for colonising sites associated with the shoreline of lakes, more hydrated and almost devoid of associated plants. For *Drosera intermedia*, it was revealed that it acclimatises to mid-lake bogs rich in magnesium, nitrogen, phosphorus and manganese. In addition, it has been shown that an important reason for the receding of sundews may be changes associated with rising temperatures in the study area.

**Keywords:** *Drosera intermedia*; *Drosera anglica*; environmental factors; conservation actions; cluster analysis; principal component analysis.

## Some properties of multi-way block designs

IDZI SIATKOWSKI<sup>1\*</sup>

<sup>1</sup>Department of Mathematical and Statistical Methods  
Poznan University of Life Sciences  
<sup>\*</sup>idzi.siatkowski@up.poznan.pl

We consider the decomposability property and the connectedness of the multi-way block design. Further, we characterize the efficiency-balance and the variance-balance of a class of multi-way block designs in relation to the corresponding balance properties of subdesigns.

## **Application of multivariate statistical methods in the prediction of detecting early phosphorus deficiency in tomato plants**

LESZEK SIECZKO<sup>1\*</sup>, KATARZYNA KOWALCZYK<sup>2</sup>, JANINA GAJC-WOLSKA<sup>2</sup>,  
MOHAMED HAZEM KALAJI<sup>3</sup>, PIOTR DĄBROWSKI<sup>4</sup>

<sup>1</sup>Institute of Agriculture

<sup>2</sup> Institute of Horticultural Sciences

<sup>3</sup> Institute of Biology

<sup>4</sup> Institute of Environmental Engineering

<sup>1, 2, 3, 4</sup>Warsaw University of Life Sciences in Warsaw, Poland

\*Leszek\_Sieczko@sggw.edu.pl

Nutrients deficiency in plants is one of the main factors adversely affecting plant growth and yield of agricultural and horticultural production. Usually detection of nutrients deficiency in plants is based on visible symptoms or destructive chemical analysis. Chlorophyll a fluorescence is one of the most used approaches to investigate photosynthesis light-dependent reactions. It is widely accepted that it can inform about structural and functional changes in photosynthetic apparatus, especially under abiotic stresses. The aim of this research was to study the influence of phosphorus deficiency on photosynthetic efficiency of tomato plants, and to assess the potential of chlorophyll a fluorescence measurements in early detection of this stress. The changes in photosynthetic machinery performance were observed earlier than the visible effects of phosphorus deficiency. Some sensitive parameters of OJIP test were altered but we were not able to identify the phosphorus deficiency stress based on the alteration of only one of these parameters. We conclude that the detection of specific stress reactions of plants to phosphorus deficiency should be based on integrative analysis of the response of the photosynthetic machinery. Such stress induced reactions can be detected by Principal Component Analysis of a large number of JIP-test parameters.

Keywords: chlorophyll fluorescence, tomato, phosphorus deficiency, principal component analysis

## Bayesian modelling of variance components makes analysis of resolvable incomplete block designs more efficient

MARCIN STUDNICKI<sup>1\*</sup>, HANS PETER PIEPHO<sup>2</sup>

<sup>1</sup> Department of Biometry; Institute of Agriculture, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warsaw, Poland;

<sup>2</sup> Biostatistics Unit, University of Hohenheim, Fruwirthstraße 23, 70599 Stuttgart, Germany

\*[marcin\\_studnicki@sggw.edu.pl](mailto:marcin_studnicki@sggw.edu.pl)

Nowadays, the standard approach to variance components estimation in LMM for METs is the restricted maximum likelihood (REML) method (Kleinknecht et al., 2013). One drawback of the REML method in the context of incomplete block designs is that estimates of the block variance may be estimated as zero (Sarholz and Piepho, 2008), which can compromise the recovery of inter-block information and hence reduce the accuracy of Best Linear Unbiased Predictions (BLUPs) and Best linear Unbiased Estimators (BLUE) of treatment effects. This, in turn, may affect the reliability of cultivar evaluation and the credibility of recommendations to farmers. Due to the development of statistical and computational methods, there is an increasing interest in adopting Bayesian approaches to METs data analysis (Crossa et al., 2011). One of the advantages of the Bayesian approach is the possibility to define a prior probability distribution for parameters of interest. This allows taking into account the knowledge about parameters available before a new experiment. A prior distribution can be based on information available for the cultivars, its pedigree and trial locations. The incorporation of this information into the model allows improving the analysis of new experiments, especially when involving highly unbalanced datasets. The use of such additional information in cultivars evaluation and METs analysis is very rare. In order to increase the precision of the analysis of individual trials laid out as alpha designs in MET, we here make a proposal to create a prior distribution for variance components for replicates, blocks and plots, based on the results of previous (historical) trials. We propose different modelling approaches for the prior distributions and evaluate the effectiveness of the Bayesian approach to the REML method classically used with MET. The historical data set come from real field trials with wheat of the Polish Post-Registration Variety Testing System (PVTs), an official cultivar testing system used for informing recommendations to farmers. For individual trials laid out as alpha-designs we assume the standard linear model. Our modelling approach is based on sums of squares of a sequential analysis of variance (ANOVA) for this model computed for the individual trials (Sarholz and Piepho, 2008). Our approach can be regarded as a Bayesian extension of classical ANOVA estimation of variance components (Searle et al., 1992). We compute the sums of squares and their expected values using the MIXED procedure in SAS. Conditionally on the variance components,  $\sigma = (\sigma_r^2, \sigma_b^2, \sigma_e^2)^T$ , the sums of squares have scaled central chi-squared distributions. For the variance component vector, we assume four different prior specifications. The first is a trivariate log-normal (Sarholz and Piepho, 2008) distribution. In our second specification, to ensure that the parameters of the variance-covariance matrix for the log-variances are properly constrained to suitable values. The exponential specification for the variances makes sure these are positive, whereas the use of the inverse of Fisher's z transformation ensures that a correlation  $\pi$  obeys the constraint. This second specification has the same set of hyper-parameters as the first specification for the trivariate log-normal distribution but in a differently parameterized form. We also consider other prior distributions for the variances. When implementing these in the NLMIXED procedure of SAS, we faced the challenge that these only allow normally distributed random effects. This works fine for our two first specifications involving the trivariate normal,

but requires some modification for other prior distributions. Specifically, we use the Gamma and inverse Gamma distributions for the variances. To fit these, we use the inverse distribution function method [also known as inverse cumulative distribution function (c.d.f.) method]. For a full Bayesian approach we used the Markov Chain Monte Carlo (MCMC) method with Gibbs sampling in the MCMC procedure for SAS. Additionally, in the full Bayesian approach, for the vector of variance component means we assumed multivariate normal hyperprior distributions with mean vector  $\mu=0$  and a variance matrix  $V = 1e10*3$ . We used an inverse-Wishart prior distribution,  $W^{-1}(v;\Omega)$ , for the variance-covariance matrix of the log-variance components, where  $v$  is the degrees of freedom equal to the number of variance components and is a scale matrix as diagonal matrix with values 1. The hyper-parameters  $\alpha$  and  $\beta$  in the gamma or inverse gamma prior distributions were estimated using the maximum likelihood ML methods (Bar and Schifano, 2011). This is expected to be particularly worthwhile for the block variance in case of a limited number of blocks and small block size, in which case REML estimates frequently converge to zero thus effectively reducing the model by dropping the block effect and foregoing any within-replicate adjustments. Our rationale here is that incomplete blocks are always expected to capture some of the within-replicate heterogeneity and hence an estimate of exactly zero is not usually plausible. A similar reasoning applies to the variance for replicates, even though there is not usually any inter-replicate information to be recovered. This reasoning leads to our expectation that a Bayesian approach can improve the mixed model analysis of individual trials. Our suggested approach is to simply plug in the variance component estimates coming out of our hierarchical models into the mixed model package, which is then used to solve the mixed model equations. The purpose of the simulation is to test our hypothesis that the Bayesian approaches can improve the recovery of inter-block information and hence lead to more accurate estimates of treatment means and their differences. The simulation study is based on the historical data set from Polish Post-Registration Variety Testing System. On this basis, we simulated individual trials ( $n = 1000$ ). We used two types of simulated data sets, one with a small number of blocks (2) and small size of blocks (3 plot per block), and one with a relatively large number (10 blocks) and size of blocks (10 plot per block). For an each simulated individual trial we determined the mean squared error of estimated treatment (cultivars) differences (MSED) for all possible treatment means (cultivar means) pairwise comparison. This was later averaged for the all simulated trials separately for each model. The lower the value of the MSED, the more efficient the recovery inter-block information.. Additionally, for actual variances for new individual trials Bayesian estimated based on the “known” prior were used to evaluate the prediction of estimated variance components using the mean squared error (MSE)

#### References

- Bar, H.Y., and E.D. Schifano. 2011. Empirical and fully Bayesian approaches for random effects models in microarray data analysis. *Stat. Modelling* 11(1): 71–88. doi: 10.1177/1471082X1001100105.
- Crossa, J., S. Perez-Elizalde, D. Jarquin, J.M. Cotes, K. Viele, et al. 2011. Bayesian Estimation of the Additive Main Effects and Multiplicative Interaction Model. *Crop Sci.* 51(4): 1458–1469. doi: 10.2135/cropsci2010.06.0343.
- Jarquín, D., S. Pérez-Elizalde, J. Burgueño, and J. Crossa. 2016. A Hierarchical Bayesian Estimation Model for Multienvironment Plant Breeding Trials in Successive Years. *Crop Sci.* 56(5): 2260–2276. doi: 10.2135/cropsci2015.08.0475.
- Kleinknecht, K., J. Möhring, K.P. Singh, P.H. Zaidi, G.N. Atlin, et al. 2013. Comparison of the Performance of Best Linear Unbiased Estimation and Best Linear Unbiased Prediction of Genotype Effects from Zoned Indian Maize Data. *Crop Sci.* 53(4): 1384. doi: 10.2135/cropsci2013.02.0073.
- Sarholz, B., and H.-P. Piepho. 2008. Variance Component Estimation for Mixed Model Analysis of cDNA Microarray Data. *Biometrical J.* 50(6): 927–939. doi: 10.1002/bimj.200810476.



# New members of the Johnson family of probability distributions: properties and application

PIOTR SULEWSKI<sup>1\*</sup>

<sup>1</sup>Institut of Exact and Technical Sciences  
Pomeranian University in Słupsk, Poland  
[\\*piotr.sulewski@apsl.edu.pl](mailto:piotr.sulewski@apsl.edu.pl)

In this paper we introduce new non-mixed bimodal distributions belonging to the Johnson family of distributions (JFD) named SC and SD, where SC is a special case of SD. The SD is the only distribution among the JFD that can be both unimodal and bimodal distribution. Properties of the SD are methodically studied. The SC and SD are compared with the seven (except the normal distribution) members of the JFD for flexibility and applicability. In order to test for flexibility, a special measure called skewness-kurtosis-square is defined. The best dispersion of points with coordinates (skewness, kurtosis) occurs for the SD and for the well-known SU and SB. Two real datasets were used to test the applicability. EN turned out to be better than its competitors in terms of information criteria and results of three goodness-of-fit tests. The content of the paper shows that the EN (including SC) as a new member of the Johnson family of distributions and simultaneously as a new distribution from the non-mixed bimodal distribution category, is a competitive model that deserves to be added to the existing distributions in modeling data.

Keywords: departure from normality, flexibility of distribution, normal distribution

## References

- Balakrishnan, N., Ristic, M.M. '2016. *Multivariate families of gamma-generated distributions with finite or infinite support above or below the diagonal*. Journal of Multivariate Analysis, 143: 194–207.
- Johnson, N.L. 1949. *System of frequency curves generated by methods of translation*. Biometrika 36: 149–176.
- Sulewski, P. 2019. *Two-Piece Power Normal Distribution*. Communications in Statistics – Theory and Method, DOI:10.1080/03610926.2019.1674871.
- Sulewski, P. 2019. *Modified Lilliefors goodness-of-fit test for normality*. Communications in Statistics – Theory and Method, DOI:10.1080/03610918.2019.1664580.
- Sulewski, P. 2020. *Normal distribution with plasticizing component*. Communications in Statistics – Theory and Method, DOI:10.1080/03610926.2020.1837881.

## The use of Rao's score test in the performance of experiments.

ANNA SZCZEPAŃSKA-ÁLVAREZ<sup>1\*</sup>

<sup>1</sup>Department of Mathematical and Statistical Methods  
Poznan University of Life Sciences in Poznań, Poland  
<sup>\*</sup>anna.szczepanska-alvarez@up.poznan.pl

The way an experiment is conducted affects its analysis and results. In this presentation the Rao's score test is used to check a contamination of compounds during the performance of the experiment. The three level model with  $\Sigma \otimes \Psi_1 \otimes \Psi_2$  covariance structure, where  $\Sigma$  is an arbitrary positive definite matrix, and  $\Psi_1$  and  $\Psi_2$  both are correlation matrices with the compound symmetric structure, is considered. It is worth to notice that even for the simple assumption about the same correlation between levels of factors, the inference should be carried off carefully. To illustrate the results a fermentation process is presented.

Keywords: maximum likelihood estimation, Rao's score test, three level model

### References

- Filipiak K., Klein D., Roy A. 2016. *Score test for a separable covariance structure with the first component as compound symmetric correlation matrix*. Journal of Multivariate Analysis 150: 105–124.
- Kollo T., von Rosen D. 2005. *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht.
- Rao C.R. 2005. *Score Test: Historical Review and Recent Developments*. In: Balakrishnan N., Kannan N., Nagaraja H.N. (ed) *Advances in I; Methodology and Application*. Birkhauser, Boston, pp 3-15.
- Szczepańska-Álvarez A., Álvarez A., Szwengiel A., von Rosen D. *Independence factor levels in a three-level model*. In preparation.

## **Wpływ środowiska glebowego na kształtowanie się cech dendrometrycznych jesionu wyniosłego (*Fraxinus excelsior* L.) w aspekcie zagrożeń biotycznych i abiotycznych gatunku**

KRZYSZTOF TURCZAŃSKI<sup>1</sup>, KATARZYNA KAŻMIERCZAK<sup>2</sup>, BOGNA ZAWIEJA<sup>3\*</sup>

<sup>1</sup>Katedra Botaniki i Siedliskoznawstwa Leśnego, Uniwersytet Przyrodniczy w Poznaniu

<sup>2</sup>Katedra Urządzania Lasu, Uniwersytet Przyrodniczy w Poznaniu

<sup>3</sup>Katedra Metod Matematycznych i Statystycznych, Uniwersytet Przyrodniczy w Poznaniu

\*[bogna.zawieja@up.poznan.pl](mailto:bogna.zawieja@up.poznan.pl)

Jesion wyniosły (*Fraxinus excelsior* L.) występuje niemal w całej Europie. Jako klimaksowy składnik lasów pojawia się w środkowym pasie kontynentu, tworząc cenne siedliska przyrodnicze. Niestety, od 1992 roku udział drzewostanów jesionowych stopniowo maleje. Głównymi czynnikami są niekorzystne warunki biotyczne (grzyb *Hymenoscyphus fraxineus*) oraz abiotyczne (spadek poziomu wód gruntowych). Zamieranie gatunku najbardziej widoczne jest na siedliskach optymalnych tj. cechujących się wysoką wilgotnością i żyznością. Obecnie prowadzone są prace nad możliwością przetrwania jesionu na siedliskach mniej optymalnych o mniejszej żyzności i wilgotności. Dotyczy to zwłaszcza lasów świeżych (Mebf), w których proces zamierania jest mniej nasilony. W związku z tym, celem niniejszych badań jest porównanie cech dendrometrycznych jesionu tj. pierśnicy i wysokości na siedliskach nietypowych dla gatunku (Mebf) z siedliskami optymalnymi (ols jesionowy, Aaf; las łęgowy, Ff; las wilgotny, Mbf) oraz określenie wpływu parametrów glebowych na kształtowanie się wspomnianych cech dendrometrycznych. Do badań wytypowano 24 drzewostany na siedliskach Aaf i Ff (łącznie 8), Mbf (8) i Mebf (8). W każdym drzewostanie założono powierzchnię badawczą wielkości 400 m<sup>2</sup>, na której pomierzono pierśnice i wysokości 15 jesionów należących do I i II klasy Krafta. Następnie w centralnej części powierzchni wykonano opis profilu glebowego wg klasyfikacji IUSS WRB (2015), określono poziom wody gruntowej i pobrano próbki do analiz fizyko-chemicznych gleby z dwóch głębokości tj. 20 i 60 cm (pH gleby, gatunek gleby, zawartość CaCO<sup>3</sup>, materii organicznej (SOM), Corg, N, Ca, K, Na, Mg, Fe, Mn i stosunek C/N). Analiza statystyczna została podzielona na trzy etapy tj.: sprawdzenie czy typy siedlisk i grupy gleb się różnią od siebie; analizę kształtowania się pierśnicy i wysokości w zależności od siedliska i gleby; określenie wpływu parametrów gleby na kształtowanie się cech dendrometrycznych jesionu. Metody statystyczne objęły analizę wariancji, analizę regresji liniowej i regresji wielokrotnej (w przypadku niespełnienia założeń stosowano permutacyjne wersje tych analiz). Ponadto zastosowano niemetryczne skalowanie wielowymiarowe oraz analizę korelacji kanonicznych. Przed przystąpieniem do analiz wykluczono wpływ wieku drzew na cechy dendrometryczne stosując analizę regresji liniowej. Jako nową zmienną, niezależną od wieku, zastosowano poprawione wartości cechy uzyskane poprzez dodanie do średniej reszt z modelu. Analizy wykazały, iż na poziomie istotności 0.05 typy lasów i grupy gleb różnią się istotnie pod względem obu analizowanych cech dendrometrycznych drzew oraz pod względem cech gleby. Wykazano ponadto istotne różnice pomiędzy typami lasu pod względem większości parametrów gleby. Stwierdzono, iż grupy gleb mają istotny wpływ na parametry drzew. Rozpatrując pierśnicę drzew wyróżniono dwie grupy. Pierwszą stanowią gleby siedlisk świeżych tj. Arenosols i Cambisols (drzewa o mniejszej pierśnicy), drugą gleby siedlisk optymalnych - Fluvisols, Gleysols, Luvisols i Phaeozems. Analiza wysokości pozwoliła stwierdzić, iż na siedliskach z glebami Cambisols drzewa były istotnie niższe.

Analiza parametrów gleb na cechy drzew, wykonana metodą regresji liniowej, wykazała, iż na wysokość nie miały istotnego wpływu następujące parametry gleb:  $\text{CaCO}_3$  20 cm, C/N 20 i 60 cm, Mn 20 cm, zaś na pierśnicę C/N 20 i 60 cm, K 60 cm, Na 20 cm i Mg 20 cm. Oprócz K 60 cm wszystkie te parametry istotnie różnicowały typy gleb, jednak nie wpływały na parametry drzew. Analiza korelacji kanonicznych pomiędzy parametrami drzew a nieskorelowanymi parametrami gleby wykazała silną, choć nieistotną zależność (współczynnik korelacji kanonicznej  $R = 0,9215$  był wysoki choć nieistotny ( $p = 0.0686$ )).

## Yield progress in maize over the period 1996–2020 based on the VCU trial data from the Czech Republic

MARTIN TLÁSKAL<sup>1\*</sup>, JIŘÍ HARTMANN<sup>1</sup>

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

\*martin.tlaskal@ukzuz.cz

Maize together with wheat, rice and soybean belongs to the top four major crops. To meet demands of increasing human population, higher yields for a unit area are desirable. This goal can be achieved by successful breeding new varieties with higher yield potential. Apart from the genotype, the yield potential of a crop is also influenced by other factors like e.g. climate conditions and soil properties. We analyzed 25 years of forage maize (*Zea mays* L.) yield data from Value for Cultivation and Use (VCU) trials to determine genetic and non-genetic trends of yields in time. For this purpose we applied a linear mixed model with regression terms. Using this model we estimated a genetic trend of 0.112–0.138 ton ha<sup>-1</sup> year<sup>-1</sup> for grain maize and 0.147–0.159 ton ha<sup>-1</sup> year<sup>-1</sup> for silage maize.

Keywords: multienvironment trials, yield progress, linear mixed model

### References

- Laidig F., Piepho H. P., Drobek T., Meyer U. 2014. *Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends*. Theoretical and Applied Genetics 127 (12): 2599–2617.
- Piepho H. P., Laidig F., Drobek T., Meyer U. 2014. *Dissecting genetic and non-genetic sources of long-term yield trend in German official variety trials*. Theoretical and Applied Genetics 127 (5): 1009–1018.
- Schils R. L. M., Van den Berg W., Van der Schoot J. R., Groten J. A. M., Rijk B., Van de Ven G. W. J., Van Ittersum M. K. 2020. *Disentangling genetic and non-genetic components of yield trends of Dutch forage crops in the Netherlands*. Field Crops Research 249, 107755.

## **Taguchi method under agricultural field experiment conditions**

DARIUSZ ZAŁUSKI\*, MARIUSZ STOLARSKI, MICHAŁ KRZYŻANIAK

Department of Genetics, Plant Breeding and Bioresource Engineering  
University of Warmia and Mazury in Olsztyn, Poland

\*dariusz.zaluski@uwm.edu.pl

The aim of this study was to present and evaluate the applicability of the Taguchi method in short rotation willow coppices in long-term field experiments under environmental stress. This is the first ever study where the Taguchi method was deployed in an agricultural field experiment to identify the levels of two fundamental production factors (genotype and planting density) that were most effective in stabilizing willow (*Salix* spp. L) biomass yields under variable uncontrollable abiotic and biotic stresses in annual harvest rotation. The analysis was based on the results noted in the first 12 consecutive years of a long-term field experiment in northern Poland. The research assumptions were validated, which suggests that the Taguchi method could be considered in agricultural field experiments.

**Keywords:** experimental design; long-term field experiment; *Salix*

### **Acknowledgements**

The results presented in this paper were obtained as part of a comprehensive study financed by the University of Warmia and Mazury in Olsztyn, Faculty of Agriculture and Forestry, Department of Genetics, Plant Breeding and Bioresource Engineering (grant No. 30.610.007-110).

## Confidence interval for the fraction of sensitive questions

WOJCIECH ZIELIŃSKI<sup>1\*</sup>

<sup>1</sup>Katedra Ekonometrii i Statystyki

SGGW w Warszawie

\*wojciech\_zielinski@sggw.edu.pl

In socioeconomic research estimation of the fraction of positive answers to a sensitive question. The sensitive question is such that the respondent would not give a honest answer. In the paper the construction of a confidence interval for such fraction.

Keywords: sensitive questions, model of nonrandomized responses, confidence interval