DESIGN AND EVALUATION OF EXPERIMENTS WITH SAS®

Adriana Horníková [1]

[1] University of Economics in Bratislava (Faculty of Economics Informatics, Institute for Statistics), Dolnozemská cesta 1/b, 853 25 Bratislava, Slovakia, e-mail: ahornik@euba.sk

Abstract − Design of experiments is a theory that can not be neglected in any technical education as well as in applied statistics or economics education. For longer than a year lasted a project aiming to publish a new book in Slovak language entitled Design of Experiments with SAS. This book of eleven Chapters combines theoretical knowledge in the scientific area of design and evaluation of experimental results with application of SAS® statistical software.

A model of a simple strategy for design of experiments starts with description of a problem, choosing an appropriate design with responses and factors, collection of experimental data which is followed by fitting of an appropriate model that can be used for predicting.

Keywords: Design of Experiments, Evaluation of Measurements, SAS Statistical Software.

1. INTRODUCTION TO DESIGNING WITH SAS

In winter term 2008/09 was in Slovakia at the University of Economics in Bratislava developed and taught a new lecture Design of Experiments and their Evaluation (abr. NAVEX). For students support, but also for other interested readers and scientists a new book with the same title was prepared and will be published. To enhance knowledge of this book practical applications with SAS statistical software were selected to support numerical examples.

The whole theoretical and practical knowledge of the book is given as a compact approach. Starting with a short introduction detailing how is defined the design of experiments through out various literature and giving the engineering strategy on solving problems by making use of the design of experiments approach [3]. All definitions throughout the entire book are in compliance with the valid technical standard ISO 3534 part 3 [1].

SAS system is a typical modular system with several possibilities for designing an experiment. Design dedicated module of 12 platforms is called SAS JMP® (Fig. 1.). It enables data analysis of various kinds of variables, like ordinary, continuous up to quality control statistical tools, e.g. Shewhards´ diagrams, etc. One platform is dedicated to creating and evaluation of different types of designs of experiments. This design platform comprises of 10 sub-platforms, which are:

- screening design platform suitable for the first start with a design problem,
- custom design platform suitable for creating designs with specifications,
- response surface design platform for creating designs with models of higher order response surface terms,
- nonlinear design platform for non linear problems,
- space filling design platform with special functions,
- full factorial design platform for factorial designs,
- Taguchi design platform for designs with inner and outer arrays,
- mixture design platform for designs with mixture requirements,
- augment design platform suitable to add centre and axial points to a design or to fold over a design, etc.
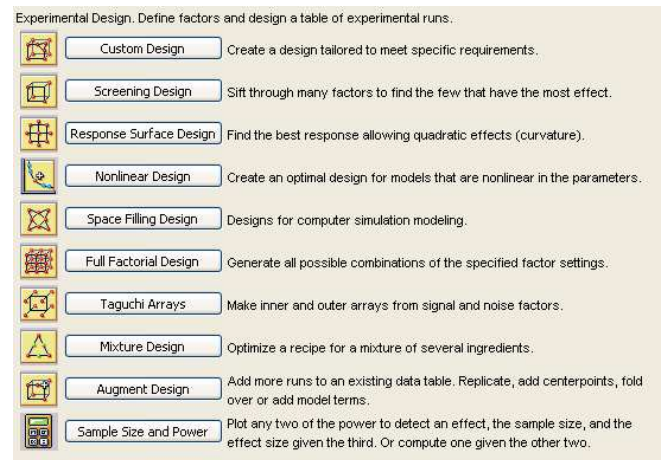- sample size and power platform [9].

Figure 1. SAS JMP® system, separate module - starter window.

Separate functions can be also found in SAS Enterprise Guide, mainly functions for one way analysis of variance (ANOVA) with a not parametric alteration (Kruskal-Wallis test) and generalized linear models fitting (function GLM). For analysis of data sets not meeting the homoscedasticity requirement is available the Welch´s variance weighted ANOVA test.

Another module incorporated in the SAS Foundation is the ADX Interface (Fig. 2). It has enhanced capabilities for designing experiments and comes with the SAS QC (Quality Control) module. For programming interested, QC module has functions for quality control and other statistical technical applications and computations as well as four

functions relating to statistical design of experiments (PLAN, FACTEX, OPTEX, ANOM) [6-8].

Figure 2. The ADX Interface starting window with data.

2. STEPS FOR DESIGNING AN EXPERIMENT

The overall engineering strategy of using the theory of design of experiments can be very simplified described in the following five steps. First comes the description of the problem under consideration, followed by designing a comprehensive experiment, realization of it and then based on creating a suitable statistical model with appropriate analysis and the analyzed problem evaluation and interpretation represent the solution. Very often not just statistical computational methods, but also graphical representation is used for data evaluation. Next a prediction for a new observation or treatment can be made and experimentally proven.

Figure 3. Screening design platform, input window for defining responses and factors accompanied with designs´ proposal in window part Choose a Design.

Not loosing on generality the design of experiments can be described in several steps:
- selection of responses, factors, their kinds and levels, blocks, defining experimental space and constrains;
- determine interaction terms, sample size, repeats, qualities of the proposed design (orthogonality, rotatability, optimality, etc.), see Figure 3;
- gather experimental data and analysis of results, if needed predict [9].

System SAS JMP® provides the user with different types of factors. They can be continuous that is that they are exclusively numeric data types. Continuous factors have values defined by lower and upper limits. Categorical factors can be either numeric or character that is ordered by numeric magnitude or sorting sequence. Blocking factors are similar with categorical factors, but the number of runs is limited. Covariate factors are not controllable, but their values are known in advance. Mixture factors are continuous factors that are ingredients in a mixture. Constant factors do not change their values during the experiment. Uncontrolled factors have values that can not be controlled during the experiment, but in the same time these are factors that should be included in the model [3, 9].

2.1. One factor designs

First Chapter is dedicated to one factor experiments. Introducing analysis of variance with several tests, like equality of variances test (homoscedasticity test), normal distribution requirement proof, testing of residuals up to multiple comparison after rejecting the null hypothesis of equality of the response mean values for all levels of the tested factor. More important from the point of view of the design is to properly distinguish between fixed and random factors. Such definition of a factor is based on the technical knowledge or experimental experiences.

In general after the decision on kind of used factor in the design was made, the appropriate model is mostly given. Throughout the entire book are used linear additive models. The model of one factor designs consists of a general mean value which is constant (μ) plus one factor called A and the independent random errors (ε_{ij}). All these terms added together build the parametric model of response, here y_{ij} in (1), when the requirement of normal distribution of random errors in (2) and the homoscedasticity of variances in (3) are met, as follows

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \tag{1}$$

when

$$\varepsilon_{ij} \sim N\left(0, \sigma_e^2\right) \tag{2}$$

Factor itself can be of two kinds. They can be fixed, so called fixed factors. Lets assume a balanced one factor experiment (3)

$$n_1 = n_2 = \ldots = n_a = n \tag{3}$$

with n_i being the sample subset size for a given level of the fixed factor A. In this case the factor A is defined with a

different fixed levels ($i = 1, 2, \ldots, a$) that are designated in the model (1) as α_i. An example of a fixed levels factor could be a factor defining the concentration while producing paper and the response variable is the strength of the paper.

If response is statistically significantly different for different levels of factor A, we can declare such factor A being influential on the response. The statistical hypothesis testing to decide on this issue is based on the null hypothesis that assumes equal influences at different factor levels on the response. For a factor with fixed levels, this test would be as follows

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_a$$

$$H_1 : \alpha_i \neq \alpha_j \text{ at least for one pair of } i \neq j \qquad (4)$$

This technique, testing according (4), is called multiple comparison test. Its aim is to prove different response variable values at different factor levels of factor A. There are several methodologies how to approach this problem. A simpler strategy is to compare pairs of means of subsets and decide based upon the result (e.g. Each Pair Student´s test). Slightly different situation is when the means of subsets have to be compared with a given constant (e.g. Dunnet´s multiple comparison test). More sophisticated approach is to compare all means of subset at once by linear equations. In this case we speak about linear contrasts (All Pairs Tukey-Kramer test or Best Hsu´s MCB test). These results can be easily graphically presented in the form of diamonds graph or circles diagram for multiple comparison tests, as is shown in fig. 4, using different colours and thickness of the line for distinguishing between groups as well as groups against the nominal value.
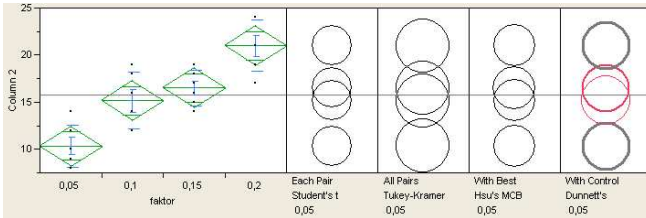

Figure 4. Multiple comparison test with fixed factor of four levels.

Every subset mean value follows normal distribution (5)

$$\bar{y}_{i.} \sim N\left(\mu_i, \frac{\sigma^2}{n}\right) \qquad (5)$$

where $\bar{y}_{i.}$ is the average of all observations in the ith group and the variance σ^2 is estimated using the MS_E statistic. The model (1) assumes also equal variances in subsets, these are build based on the equal factor level. To test the equality of variances (6), this quality is called homoscedasticity use

$$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_a^2 \qquad (6)$$

The SAS JMP system offers three tests, the Bartlett´s test, Brown-Forsythe´s test and Levene´s test.

The parameters estimation in the model with one fixed factor is limited. Generally there is no unbiased estimate of the overall mean value μ neither of the ith-factor level α_i, but a simple approximation uses the overall mean

$$\hat{\mu} = \bar{y}_{..} \qquad (7)$$

An estimate of the difference at two different factor levels is

$$\hat{\alpha}_i - \hat{\alpha}_j = \bar{y}_{i.} - \bar{y}_{..} \text{ for } i = 1, 2, \ldots, a \qquad (8)$$

Very often is part of the evaluation estimation a confidence interval (9) for the subset mean value on the ith level of the fixed factor with the $100(1-\alpha)$ % confidence

$$\bar{y}_{i.} \pm t_{1-\alpha/2}(N-a) \cdot \sqrt{\frac{MS_E}{n}} \qquad (9)$$

where $t_{1-\alpha/2}(N-a)$ is the $(1-\alpha/2)$-quatile of t-Student´s distribution with $N-a$ degrees of freedom and N is the number of all observations, here equals to $N = an$.

The overall model significance is tested based on the comparison of the model explained variability with the residual variability. These variabilities are independent and add up to the total variability of the response (10), here the variability expressed in the form of sum of square (abr. SS)

$$SS_T = SS_A + SS_E \qquad (10)$$

while the test statistic (11) is based on the two sample variance equality test

$$F_{test} = \frac{\dfrac{SS_A}{(a-1)}}{\dfrac{SS_E}{N-a}} = \frac{MS_A}{MS_E} \qquad (11)$$

where the null hypothesis states that the whole model is insignificant and the alternative hypothesis defines the model as significant. If the inequality in (12) is true, the null hypothesis is being rejected and the model by ANOVA test is declared significant

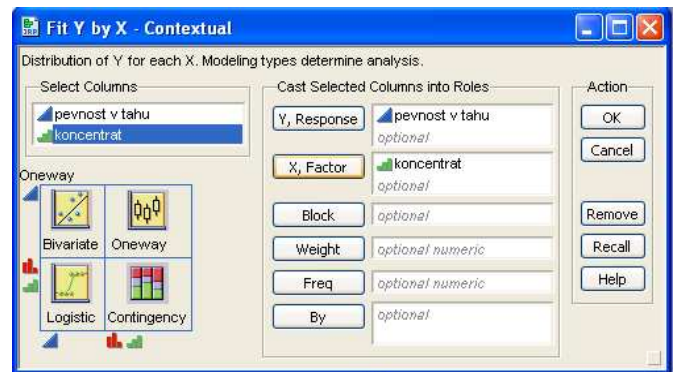$$F_{test} > F_{1-\alpha}(a-1, N-a) \qquad (12)$$

Figure 5. The SAS JMP® system's one-way ANOVA window.

2.1.1. Random effects

The second possibility for the factor declaration is to be random, in which case factor B is defined by variance σ_β^2 (15). Let's assume the following statistical model (13)

$$y_{ij} = \mu + \beta_i + \varepsilon_{ij} \qquad (13)$$

and independent random errors normally distributed

$$\varepsilon_{ij} \sim N\left(0, \sigma_e^2\right) \qquad (14)$$

and random factor

$$\beta_i \sim N\left(0, \sigma_\beta^2\right) \qquad (15)$$

The variance for each observation is

$$\mathrm{Var}\left(y_{ij}\right) = \sigma_\beta^2 + \sigma^2 \qquad (16)$$

where sigmas squared in (16) are called variance components and then the overall model is called the model with variance components. Prerequisite requirements of the model (13) are

$$\mathrm{cov}\left(\beta_i, \beta_j\right) = 0 \text{ for all } i \neq j$$

$$\mathrm{cov}\left(\beta_i, \varepsilon_{jk}\right) = 0 \text{ for all } i, j, k$$

$$\mathrm{cov}\left(\varepsilon_{ij}, \varepsilon_{kl}\right) = 0 \text{ for all } i \neq k$$

$$\mathrm{cov}\left(\varepsilon_{ij}, \varepsilon_{il}\right) = 0 \text{ for all } j \neq l \qquad (17)$$

An example of a random effect is when different children have different starting point, when approaching to study their reaction. Then testing the significant influence of a random factor (defined in compliance with (15)) one can proceed with a hypothesis test (18) as

$$H_0 : \sigma_\beta^2 = 0$$

$$H_1 : \sigma_\beta^2 > 0 \qquad (18)$$

Results of factor significance on response from the influence test (in (4) or (18)) and the overall significance of the model itself, are summarized in ANOVA table which comprises of these columns: title of the source of variability, sum of squares, degrees of freedom, means of squares, F-test and p-value. ANOVA tables are nowadays used in this standard format. Other statistics that accompany ANOVA model are R-Square, coefficient of variation or Root MSE.

Subsequent testing of residuals for normality can be done by several statistical tests. To list several widely spread, one can mention the Shapiro-Wilk´s, the Kolmogorov-Smirnov´s, the Cramer-von Mises´s or the Anderson-Darling's tests. Graphical approaches are often used in the form of probability plot, quantile plot or histogram. In the case of residuals are not normally distributed these parametric approaches of evaluation can not be used and data has to be re-evaluated by non-parametric tests.

2.1.2. Power and sample size

The second Chapter of the monograph continues the first Chapter with computations of sample size (n), power ($1-\beta$) and significant difference (δ). Available is also a computation of sample size for detecting a change in difference at significance level (α). SAS JMP system enables plotting of operation curves for different combinations of the probability of the error of the first kind (α), the probability of the error of the second kind (β), power ($1-\beta$) and significant difference (δ). This is valid if the normal distribution assumption is met, but for any other distribution the SAS JMP system seeks a proper way to approximate it with a normally distributed random variable [9-10].

2.2. Designs with blocking factor

In the third Chapter blocking is added to designs. Simplified one can say that a factor is organized in blocks when we assume not full randomization of treatments and we introduce a block structure to the data sets observed. This is mostly captured in the model by a fixed factor without interactions with other factors within the model, even though this model could be adjusted also for interactions.

2.3. Two factor designs

Two factor designs are simple enough to explain the possibilities to combine the fixed and random factors using. Also the principles of interaction and design qualities can be presented to the reader. Three different two factor designs are presented in the monograph.

2.4. Square designs

Typical multi-factor designs with factors having several levels are lattice designs, Latin square designs, Greco-Latin square designs or Youden square designs. All these designs assume no interaction between factors. All basic designing principles like repetitions, randomization or balancedness are explained.

2.5. Multi-factor designs

Further in the subsequent Chapter is explained the full factorial design with interactions and effects estimation. Later is detailed the principle of reducing a full factorial design based on the generating relation and finding the confounding of a reduced factorial design. Design resolution principle is also explained. In this Chapter are four applications solved using the SAS JMP® system.

The model consists of a mean value which is assumed to be constant (μ), fixed factors A & B (α_i, β_j, with $i = 1,2,...,a$ and $j = 1,2,...,b$), their interactions ($\alpha\beta_{ij}$) plus independent random errors (ε_{ij}). All terms model the response, here y_{ij} in (19), when the requirement of normal distribution of random errors in (2) or (14) are met

$$y_{ij} = \mu + \alpha_i + \beta_j + ... + \alpha\beta_{ij} + ... + \varepsilon_{ij} \qquad (19)$$

One important conclusion from experimentation is whether either the main factor or interactions influence the response.

In full factorial design 2^k, a design with k fixed two level factors, is the main factor A effect λ_A estimated

$$\hat{\lambda}_A = \bar{y}_{A+} - \bar{y}_{A-} \qquad (20)$$

where $\bar{y}_{A+}$ and $\bar{y}_{A-}$ are averages of observations at factor A levels respectively. In the same way are estimated main effects of other fixed factors. The estimate of interaction between factors A & B is

$$\hat{\lambda}_{AB} = \frac{\bar{y}_{A+B+} - \bar{y}_{A-B+} - (\bar{y}_{A+B-} - \bar{y}_{A-B-})}{2} \qquad (21)$$

where $\bar{y}_{A+B+}$ is the average of observations at the factors level A+B+. Sum of squares can be easily computed from factor main effects and interactions (18) as

$$SS_A = n \cdot \lambda_A^2$$

$$SS_B = n \cdot \lambda_B^2$$

$$SS_{AB} = n \cdot \lambda_{AB}^2 \qquad (22)$$

Different multi-factor designs are detailed in different Chapters as given.

2.6. Hierarchical designs

Hierarchical designs enable estimation of variance while there are only selected levels for factor combinations. These factors´ settings do not account for interactions between factors which is a result of the hierarchical structure of factors in the selected design. These designs are often used for evaluation of laboratory experimentation where factors can be for example operator, date of experimentation, laboratory method, different machinery, etc. These designs are often used for inter-laboratory comparisons.

2.7. Response surface designs

More sophisticated design approach is to use a response surface design. These designs cover evenly the experimental space. Box-Behnken design, Doehlerts´ geometrical designs (pentagonal and hexagonal designs) and the central composite designs are also suitable for expanding the experimental space mostly by the steepest accent or the steepest descent methodology. Very often are added central and axial points that create a hypercube or a hypersphere. Rotatability and orthogonality can be features of desings. The main aim of these designs is slightly altered in comparison to classical designs as these designs provide the functional estimates of the response, in either linear form

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon \qquad (23)$$

or second order function

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum_{i<j}^{n} \sum_{j=1}^{n} \beta_{ij} x_i x_j + \varepsilon \qquad (24)$$

where the function is polynomial and β´s are coefficients and x represent the prediction variables in the model.

Second order functions are not linear any more and curvature can be calculated and tested for difference from zero.

Advantageous is to use additional central points to prove the quality of the model function established in the experiment as well as to use other types of additional points to widen the design space, e.g. axial points, etc.

2.8. Nonlinear designs

More comprehensive than response surface designs are nonlinear designs. They often model kinetics of chemical reactions that are either log normally or exponentially distributed. The SAS JMP® system uses the standard least square method to compute estimates. Another possibility is the stepwise regression method that is particularly suitable in nonlinear designs. In menu Personality of the window Fit Model with selected Model Specification are also listed Manova and Loglinear Variance approach for special problems. Another option is to use Generalized Linear Model for evaluation.

2.9. Mixture designs

Similar designs are designs for mixtures. They can be modelled in SAS JMP® by functions alike those given by (23) and (24), but also higher order functions (cubic) often called Scheffe's cubic model or altered cubic function, here in (25) a function without an intercept

$$y = \sum_{i=1}^{k} \beta_i x_i + \sum \sum_{i<j} \beta_{ij} x_i x_j + \sum \sum \sum_{i<j<k} \beta_{ijk} x_i x_j x_k \qquad (25)$$

SAS JMP® system offers users a variety of different model effects, like factorial (full factorial, factorial to a given degree, factorial sorted), response surface, mixture response surface, polynomial to a given degree, Scheffe's cubic (25) and radial.

Mixtures have to meet at least one requirement limiting the experimental space that is formulated in (26)

$$x_1 + x_2 + ... + x_k = 1 \qquad (26)$$

and that prevents in most cases from estimation of the intercept in the model function. The SAS JMP® system provides the user with this selection of mixture designs:
- simplex centroidal
- simplex lattice
- extreme vertices with the possibility of adding a linear constraint to a design
- ABCD mixture design for factor screening
- optimal design mirroring experimental needs and meet specific requirements.

2.10. Optimal designs

There are several different kinds of optimal designs, like A-, D- G- or V- optimality criterion fulfilling designs. There are two big groups of optimality criteria, one group optimizes the variance of the estimated function of

parameters, the other group optimizes the variance of the prediction (e.g. G-, V-, Q-optimality). The SAS JMP® system uses by default the most often used D-optimality criterion and also the I-optimality criterion (I- stands for integration) in conjunction with the coordinate exchange mechanism. In the SAS JMP® system's menu is also available the Bayesian D-optimality criterion. Besides the optimization function the SAS JMP® system also provides users with A-, D- and G-effectiveness values of a design and in addition with the prediction variance profile and the prediction variance surface.

2.11. Taguchi arrays

Taguchi designs use a different array structure in comparison to classical designs. These designs are composed of an inner and an outer array where inner array is a design build with controllable factors and outer array is build with noise factors. Crossing of these two arrays creates the Taguchi orthogonal design.

Another specialty of these designs is the use of the loss function and the signal to noise ratio. This approach has several advantages and was generalized from the point of view of the statistical theory. The SAS JMP® system is one of the few software packages on the market that has a Taguchi platform.

Taguchi designs are applied in three steps, first the system design (the primary design), then the design of parameters (the secondary design) and at last the tolerance design (the tertiary design). Each of these steps has its specific aim. The system has to be first designed based on the concept and the competition pressure and the price of a product put on the market. That is why Taguchi designs are so often used in manufacturing, quality improvement and statistical quality assurance. In the phase of parameter design the focus is on the improvement of the quality and performance of the product. After signal and noise factors and their influence on the overall mean value and variance of performance are identified, tolerances and factor settings can be proposed and verified in a separate experiment (confirmation treatment). This leads to quality improvement. Taguchi designs are probably the best example of practical applications of the experimental design's contribution to manufacturing improvement [1-10].

3. CONCLUSIONS

To support scientists and students a book shall be published detailing theory of Design of experiments and their evaluation and should be accompanied by real practical examples with applications to SAS® system. SAS® system has several interfaces that create or evaluate designs of experiments. The best suitable specialized statistical software for design of experiments seems to be the SAS JMP® system with a dozen of different platforms enabling the creation of various designs, e.g. classical, Taguchi, etc.

The SAS JMP® system using evaluation is based on the least squares method or stepwise regression method, both a general methodology for evaluation of experiments. Other methods are manova or log variance method, even GLM is available. Besides of other qualities the SAS JMP® system

computes a desirability function for factors and responses to find the best settings to solve an experimental problem.

3.1. Advantages of the desirability function

Big advantage of the SAS JMP® system is in determining the desirability, desirability values, desirability trace or function. This function plotted as functional relationship across factors and responses is aiming to find the most appropriate setting of factors and responses to gain the best settings. Desirability trace can be also maximalized which means that the best possible settings are found. Besides factors´ desirability functions see fig. 6, the SAS JMP® system provides users with a global desirability function (27) that is evaluated as the geometrical average of m individual desirability functions d_i
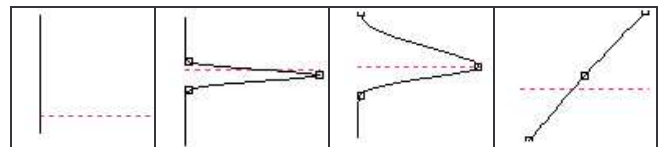
$$D = \sqrt[m]{d_1 d_2 ... d_m} \qquad (27)$$

Figure 6. Different shapes of the desirability function.

ACKNOWLEDGMENTS

REFERENCES

[1] ISO 3534-3: *Statistics. Vocabulary and symbols. Part 3: Design of experiment*, 1999.

[2] J. Goupy, L. Creighton, *Introduction to Design of Experiments with JMP Examples*. SAS Press, 2007.

[3] A. Horníková, *Design of experiments with SAS*. Iura Edition, 2009. (in print)

[4] D. C. Montgomery, *Design and Analysis of Experiments*. John Wiley and sons, 1997.

[5] R. H. Myers, D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley, New York, 1995.

[6] SAS® Institute Inc., *Getting Started with SAS 9 ADX Interface for Design of Experiments*. Cary, NC, SAS Institute Inc. 2002.

[7] SAS® Institute Inc., *SAS/QC 9 User´s Guide, Volumes 1, 2, and 3*. Cary, NC, SAS Institute Inc., 2002.

[8] SAS® Institute Inc., *System Requirements for SAS 9.1.3 Foundation for Microsoft Windows*. Cary, NC, SAS Institute Inc., 2005.

[9] SAS® Institute Inc., *JMP Design of Experiments Guide 7*. Cary, NC, SAS Institute Inc., 2007.

[10] G. Wimmer, *Biostatistics (Model with random effects)*. MATFYZ Bratislava, 1999.