ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

THE BOOTSTRAP APPROACH TO THE STATISTICAL SIGNIFICANCE OF PARAMETERS IN THE FIXED EFFECTS MODEL

Jacek Pietraszek[1], Renata Dwornicka[1], and Agnieszka Szczotok[2]

[1] Department of Software Engineering and Applied Statistics
Faculty of Mechanical Engineering
Cracow University of Technology
37 Jana Pawla II, 31-864 Krakow, Poland
e-mail: {pmpietra, dwornick}@mech.pk.edu.pl

[2] Institite of Materials Science
Faculty of Materials Science and Metallurgy
Silesian University of Technology
8 Krasinskiego, 40-019 Katowice, Poland
e-mail: agnieszka.szczotok@polsl.pl

Keywords: Bootstrap, Statistical Significance, Fixed Effects Model, ANOVA, Materials Science, Surface Layers, Nanoparticles.

Abstract. Since R.A. Fisher introduced the analysis of variance (ANOVA), the classic design of experiments (DoE) utilizes factorial model of fixed effects. The typical procedure assumes the normal distribution of the noise and uses the least square method (LSQ) to identify effects of the model with a priori assumed structure of main effects and some interactions, usually up to the second order. The terms of model are repeatedly eliminated in the specific backward stepwise regression.
Methods of approximation and prediction rapidly evolved in recent years. Apart from this, the classic approach of the fixed effects model is still very useful and popular. In fact, this model is intrinsic to ANOVA however hidden. The significance analytical improvements have appeared in the background of fixed effects models in recent decades: statistics based on widespread strict requirements of the normal distribution have been replaced by so-called robust statistics based on the weak requirement of the continuous distribution or even on the statistics based on numerical simulation like e.g. bootstrap approach. Practically, in the case of small size datasets, the conformity with the normal distribution has very weak reliability and it leads to very uncertain assessment of parameters statistical significance. The bootstrap approach with simulation-based identification of parameters confidence intervals (CI) appears to be better solution than theoretically proved but only asymptotically equal t-distribution based evaluations.

1 INTRODUCTION

The research investigation in a materials science always leads to the datasets. The large (or even huge) datasets should be analyzed to obtain summary results expressed in qualitative or quantitative forms. Such conclusions should be semi-proved by a statistical analysis.

One of the typical methodology is to observe behavior of the phenomenon against different treatments of controlled factors being precisely defined in a matrix known as a design of experiment. The general term ‘behavior’ is narrowed to a quantitative variable called ‘response’ or ‘output’. The experiment focuses on the ‘effect’: the difference of the output as a result of the difference of treatments. If only one controlled factor is changed at time, the experiment is named One-Factor-At-Time (OFAT) [1]. On the opposite side, the factorial experiment [2] may be found where all controlled factors are changed simultaneously in a specific manner. The factorial experiment is also known as ALLFAT (All-Factors-At-Time). This name is often used in an industrial environment and quality management procedures.

The sensitivity of the investigated phenomenon to changes of controlled factor decides on the statistical significance of such factor. However this definition is intuitively clear, it requires some quantitative measure and its evaluation. Typical procedure used to assessment of factors significance is the analysis of variance (ANOVA).

ANOVA [2] is a well-known procedure to identify an inequality of means evaluated for subsets selected from the general dataset by a chosen particular classification factor. The null hypothesis H_0 states that all means are mutually equal against the alternative hypothesis H_1 that not all means are equal. The ANOVA workflow leads from the general dataset through its decomposition to subsets. Mathematically, it is described as a decomposition of a variance, which finally leads to comparison of MS_{factor} (variance explained by classification factor) against MS_{error} (remained variance). Technically, it replaces the original null hypothesis to its equivalent: the ratio of MS_{factor} by MS_{error} is equal to zero. As was proved by Fisher, this ratio has F-distribution however it requires some additional assumptions about an independency of observations, a normality of residuals and a homoscedascity in subsets. These assumptions are often questionable, especially when the size of a dataset is small and the statistical inference is weak.

The proposed solution consists of two elements: (a) the replacement of the original null hypothesis with its equivalent related to parameters of an associated fixed effects model and (b) the replacement of the classic test of parameters significance with the bootstrap-based checking of zero existence inside the parameter confidence intervals. This solution is more practical for large datasets with the unknown distribution because it does not require to meet the ANOVA assumptions.

Authors have been tried [3] previously to combine a bootstrap approach with an artificial neural network approximator to analyze materials science data processed by image analysis however they have with many numerical artifacts. Such approach was more successfully for processing non-parametric statistical analysis for surface layer [4].

The bootstrap approach appears to be a promising solution for some aspects of fuzzy statistics which is developed mainly by Buckley [5, 6]. The general workflow for such implementations has been proposed by Grzegorzewski [7]. Some preliminary investigations in the field of a design of experiments (DoE) have been made by Pietraszek [8, 9].

2 COMPUTATIONAL METHODS

Two main methods were used: the fixed effects model [2] and the bootstrap method [10]. The fixed effects model is used to evaluate effects being deviations of particular means from the grand mean. The bootstrap method is used to evaluate confidence intervals for effects. Finally, the existence of zero inside the intervals is inspected. If any of intervals contains zero, the null hypothesis of ANOVA is rejected.

2.1 Fixed-effects model

The fixed-effects model [2]consist of three terms:

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

where:

μ – grand mean (average response),
α_i – an effect at *i*-th treatment,
ε_{ij} – an random error at *i*-th treatment and *j*-th replication,
y_{ij} – the observed response at *i*-th treatment and *j*-th replication.

The grand mean and effects are evaluated using the ordinary least squares (OLS) [2].

2.2 Bootstrap

The bootstrap approach [10] to fixed-effects model (Eq.1) is described by the following workflow. In the beginning, the parameters $\hat{\mu}, \hat{\alpha}_i$ of the model were identified from the dataset using the least squares criterion L

$$\min L(\mu, \alpha_i) = L(\hat{\mu}, \hat{\alpha}_i) \tag{2}$$

then fits $\hat{y}_i$ were evaluated

$$\hat{y}_i = \hat{\mu} + \hat{\alpha}_i \tag{3}$$

and at last residuals r_{ij} of the model

$$r_{ij} = y_{ij} - \hat{y}_i \ . \tag{4}$$

Then, iteratively, (a) the new dataset $\mathbf{R}_B$ of the same size containing bootstrapped residuals $r_{b|ij}$ was constructed by drawing with replacement from the set of residuals:

$$\mathbf{R}_B = \left\{ r_{b|ij} : (1 \le i \le n) \wedge (1 \le j \le r) \wedge (\forall i, j \; \exists k, l : r_{b|ij} = r_{kl}) \right\} \tag{5}$$

and (b) new bootstrapped residuals were added to the model fits creating new bootstrapped “observations”:

$$y_{B|ij} = \hat{y}_i + r_{B|ij} \ . \tag{6}$$

That new observations were used (c) to identify bootstrapped parameters of the model.

$$\min L(\mu, \alpha_i) = L(\hat{\mu}_B, \hat{\alpha}_B) \tag{7}$$

The parameters were collected until large number of iterations will be reached. Finally, (d) the quantiles were evaluated (typically 2.5% and 97.5%) for datasets of parameters as a range of confidence intervals and at last (e) zero existence inside confidence intervals was checked.

3 MATERIALS AND DATASET

The data were obtained during investigation of the ceramic shell mould of the airfoil blade casting. The description below is a brief, while details can be found in article of Szczotok *et al.* [11].

Nickel-based superalloys are mainly used in aircraft and power-generation turbines. Creep-resistant polycrystalline turbine blades are typically produced by an investment casting process. It is especially useful for making castings of complex and near-net shape geometry, where machining may be impossible or too wasteful. Studies were performed on the IN 713C superalloy. The castings described in the work were produced by the Laboratory for Aerospace Materials at Rzeszow University of Technology in Poland. One casting called GK was selected for the microstructural and statistical analysis. Final castings were cut off. The cross-sections were included and prepared as metallographic samples from nickel-based superalloy. To reveal the microstructure of the investigated material the surfaces of the samples were etched. The microstructural investigations of the cross-sections of the GK casting were carried out by a scanning electron microscope. The recorded microphotographs were next subjected to a computer-aided image analysis program to estimate quantitatively the main parameters describing the (γ+γ') eutectic islands that occurred in the investigated superalloy.

The dataset was created by counting number and size of eutectic island detected at six different traces labelled from T1 to T6 (Table 1). The issue was to check the homogeneity of the eutectic phase i.e. statistical equivalence of observations at any trace.

Trace	**No of islands**	**Mean Area**	**SE**	**Median**	**Quartiles**	
					Q1	**Q3**
T1	31	34.5	7.8	12.0	3.8	53.6
T2	49	22.0	3.8	12.1	4.2	28.1
T3	80	24.0	3.4	10.8	4.0	32.2
T4	75	26.9	3.9	14.6	4.6	35.8
T5	64	25.1	4.2	11.3	3.9	35.1
T6	61	24.5	3.5	14.2	8.6	28.4

Table 1: Descriptive statistics of raw data (source [11])

4 ANALYSIS

The typical method for checking the homogeneity of quantitative data is ANOVA. Due to the fact that area measures are positive they need to be transformed by a specific transformation into the whole real space to avoid nonsense negative values [12]. The natural logarithm was selected as transformation mapping i.e.:

$$LnY = \ln(area)\,. \tag{8}$$

Such transformation guarantees that any value has physical sense after retransformation. Descriptive statistics of transformed data are presented in Table 2.

The ANOVA protocol leads to the classic ANOVA table which decomposes the total variation into part assigned to grouping factor and the remain assigned to all other factors grouped under name 'error' which should be treated rather as 'unexplained' than only 'random error'. For the mentioned data, the ANOVA table (Table 3) revealed that critical p-Value is equal to 0.841, much greater than significance level $\alpha = 0.05$, what means that homogeneity of traces was not rejected.

Trace	No of islands	Mean LnY	SE	Median	Quartiles	
					Q1	Q3
T1	31	2.63	0.28	2.49	1.34	3.98
T2	49	2.44	0.18	2.49	1.43	3.33
T3	80	2.43	0.15	2.38	1.40	3.47
T4	75	2.55	0.15	2.68	1.53	3.58
T5	64	2.45	0.17	2.42	1.36	3.56
T6	61	2.68	0.14	2.65	2.15	3.35

Table 2: Descriptive statistics of transformed data (source [11])

Effect	SS	df	MS	F	p
Trace	3.318	5	0.664	0.411	0.841
Error	572.019	354	1.616	–	–
Total	575.338	359	1.603	–	–

Table 3: ANOVA table for transformed data (source [11])

Simultaneously, the general linear model (GLM) (Eq.1) was introduced for 6 levels (T1…T6). The identification of the model resulted in a set of parameters (Table 4). Note that lack of the parameter for T6 is a typical presentation of results by statistical programs (here: StatSoft STATISTICA v12), because all effects from T1 to T6 should sum to 0 and T6 should be deducted from such condition.

Effect	Parameter	SE	t	p [1)]	-95 CI	+95 CI
const	2.527	0.070	35.88	0.000	2.389	2.666
T1	0.100	0.199	0.50	0.617	-0.292	0.492
T2	-0.091	0.164	-0.56	0.578	-0.414	0.232
T3	-0.099	0.136	-0.73	0.467	-0.366	0.168
T4	0.021	0.139	0.15	0.880	-0.252	0.294
T5	-0.082	0.148	-0.55	0.580	-0.372	0.208

Table 4: Parameters of fixed-effects model for transformed data

[1)] p-value evaluated from inverse cumulative distribution for t-Student at d.o.f. = 354

Trace	Fit	SE	-95 CI	+95 CI	N
T1	2.627	0.228	2.178	3.076	31
T2	2.436	0.182	2.079	2.793	49
T3	2.428	0.142	2.149	2.708	80
T4	2.548	0.147	2.260	2.837	75
T5	2.445	0.159	2.133	2.758	64
T6	2.678	0.163	2.358	2.999	61

Table 5: Model fits and their confidence intervals for transformed data

The model predictions and their confidence intervals are presented in Table 5. It should be noted that such results are still from classic ANOVA for further comparisons.

Now, the approach was switched to the bootstrap. The number of draw iterations was set to 10.000 to easy selection of quantiles from the bootstrapped dataset. After the full bootstrap procedure, the descriptive statistics were evaluated for model parameters (Table 6) and model predictions (Table 7), similarly to Table 4 and Table 5.

The bounds of the confidence intervals were easy identified due to the selected number of bootstrap iterations. They were values found at positions 250 and 9750 in the sorted bootstrapped results. Similarly, the bootstrapped p-Value was evaluated as relative position of sign switching inside the sorted bootstrapped results.

Effect	Parameter mean	SE	p-Value$_{bootstrapped}$ [1)] / p-Value$_{theoretical}$ [2)]	-95 CI	+95 CI
const	2.527	0.070	0.000 / 0.000	2.388	2.665
T1	0.101	0.196	0.614 / 0.607	-0.281	0.490
T2	-0.092	0.161	0.576 / 0.568	-0.407	0.224
T3	-0.099	0.135	0.468 / 0.464	-0.364	0.167
T4	0.021	0.137	0.869 / 0.878	-0.250	0.290
T5	-0.082	0.148	0.580 / 0.580	-0.367	0.206
T6	0.151	0.147	0.310 / 0.305	-0.136	0.435

Table 6: Parameters of fixed-effects model for transformed data

[1)] bootstrapped p-Value was evaluated from relative position of sign switching inside a bootstrap table

[2)] theoretical p-Value was evaluated from ratio (mean/SE) and t-Student distribution at d.o.f. = 354

Trace	Fit	SE	-95 CI	+95 CI
T1	2.628	0.225	2.189	3.073
T2	2.435	0.179	2.090	2.785
T3	2.428	0.141	2.153	2.703
T4	2.549	0.145	2.261	2.835
T5	2.445	0.159	2.136	2.754
T6	2.678	0.160	2.360	2.988

Table 7: Boostrap fits and their confidence intervals for transformed data

The analysis of confidence intervals bounds revealed that small asymmetric exists between left and right side relative to means (Table 8). The coefficients were evaluated as studentized i.e. a quotient of a deviation and standard error, where deviation was difference between confidence interval bound and respective mean:

$$cf = \frac{\pm 95CI - mean}{SE}. \tag{9}$$

Trace / Effect	**Parameters intervals coefficients**		**Fits intervals coefficients**	
	-95 CI	**+95 CI**	**-95 CI**	**+95 CI**
const	-1.986	1.971	–	–
T1	-1.949	1.985	-1.951	1.978
T2	-1.957	1.963	-1.927	1.955
T3	-1.963	1.970	-1.950	1.950
T4	-1.978	1.964	-1.986	1.972
T5	-1.926	1.946	-1.943	1.943
T6	-1.952	1.932	-1.988	1.938

Table 8: Studentized coefficients of confidence intervals for the bootstrap model and its fits

5 CONCLUSIONS

- Bootstrap approach appears to be effective computational method to identify parameters of fixed effects model and their statistical properties.
- Bootstrap approach does not require to make *a priori* inconvenient assumptions.
- Bootstrap approach is very convenient to automatize in computational workflow and further statistical postprocessing.

REFERENCES

[1] C. Daniel, One-at-a-Time Plans. *Journal of the American Statistical Association* **68**, 353-360, 1973.

[2] O. Kempthorne, K. Hinkelmann, *Design and analysis of experiments. Vol.1. Introduction to experimental design*. John Wiley & Sons, Hoboken, NJ, USA, 2007.

[3] J. Pietraszek, A. Gadek-Moszczak, The Smooth Bootstrap Approach to the Distribution of a Shape in the Ferritic Stainless Steel AISI 434L Powders. *Solid State Phenomena* **197**, 162-167, 2013.

[4] J. Pietraszek, N. Radek, K. Bartkowiak, Advanced Statistical Refinement of Surface Layer's Discretization in the Case of Electro-Spark Deposited Carbide-Ceramic Coatings Modified by a Laser Beam. *Solid State Phenomena* **197**, 198-202, 2013.

[5] J.J. Buckley, *Fuzzy probability and statistics*. Springer Verlag, Heidelberg, 2006.

[6] J.J. Buckley, Fuzzy statistics: hypothesis testing. *Soft Computing* **9**, 512-518, 2005.

[7] P. Grzegorzewski, *Decision support under uncertainty. Statistical methods for imprecise data.* EXIT, Warszawa, 2006.

[8] J. Pietraszek, Fuzzy Regression Compared to Classical Experimental Design in the Case of Flywheel Assembly. *Lecture Notes in Artificial Intelligence* **7267**, 310-317, 2012.

[9] J. Pietraszek, A. Gadek-Moszczak, T. Torunski, Modeling of Errors Counting System for PCB Soldered in the Wave Soldering Technology. *Advanced Materials Research* **874**, 139-143, 2014.

[10] J. Shao, D. Tu, *The Jackknife and Bootstrap*. Springer, New York, 1995.

[11] A. Szczotok, J. Nawrocki, A. Gądek-Moszczak, M. Kołomycki, The Bootstrap Analysis of One-Way ANOVA Stability in the Case of the Ceramic Shell Mould of Airfoil Blade Casting. *Solid State Phenomena* **235**, 24-30, 2015.

[12] J. Pietraszek, A. Goroshko, The Heuristic Approach to the Selection of Experimental Design, Model and Valid Pre-Processing Transformation of DoE Outcome. *Advanced Materials Research* **874**, 145-149, 2014.