

THE OPTIMAL SOLUTION TO THE PROBLEMS BY REGRESSION ANALYSIS

UDC: 519.233.5
Original Scientific Paper

Branko POPOVIĆ¹

¹University of Belgrade, Mechanical Faculty, 11000 Beograd, Kraljice Marije 16, Republic of Serbia
E-mail: branko@popovic.org

Paper received: 02.03.2019.; Paper accepted: 16.04.2019.

Problems are various possible and impossible questions, tasks and riddles that are being set up and attempt to solve in different ways. However, assuming that half of all problems are not solvable and that the remaining one-fourth cannot be defined, there are good chances for a successful mathematical solution to the second quarter of the problem. Problem-solving using regression analyses allows determining the relationship between the dependent response and the independent factors through defined regression functions, relationship estimates, predicting results, simulating data, optimizing the sizes, and improving the resolution process. Although there are many regression analyses, linear regression analyses are usually used. The author proposes the application of Procedure for selecting a significant regression analysis between 16 different regression analyses and Procedures for solving regression analyses. Using this approach in this release, an optimal solution of the problem of finding the optimal best sun exposure position for the solar collector is shown.

Keywords: Regression analyses. Optimal solution for problems.

INTRODUCTION

Business management system organizations apply a set of coordinated activities to manage and manage an organization that has a variety of business, technology and service subsystems that contain a variety of processes in which certain problems arise. Processes are sets of bound potentials and activities, which convert input data into outputs in business, technology and service subsystems (Popović, 2017). Processes of business subsystems (administration, sector, bureau) create complex problems in managing the organization and managing the quality of the process results. In the process of technological subsystems (enterprises plants, workshops), various problems arise in the realization of technologies (production and servicing) of the results of the process and the installation or installation of the results of the process with the users.

In the processes of service subsystems (Services) there are problems in: researching market needs for process results, design or creation and

development of process results, planning and preparation of process performance results, recruitment and training of personnel, procurement of materials and equipment, controlling and ensuring the quality of process results, storing process results in warehouses, selling and distributing process results, servicing or maintenance the results of the process, financial and accounting affairs and securing the assets of the economic system (Popović, 2019; Popović, 2017; Fox, 1997).

Regression analysis is a set of statistical processes in statistical modeling, which enables the determination of the relationship between a quantitative or qualitative dependent variable size (Response) and one or more independent linear or nonlinear variable (Factors) through: the defined regression function, the estimates of the forms of these relations, the prediction future results, data simulation, optimization of the sizes and improvement of the problem solving process (Popović, 2019, Kirchmer, 2011).

METHODOLOGY

The basis of the regression analysis was set by 1805 French mathematician A. M. Legendre (1752-1833) with the Smallest Square Method and 1809 German mathematician C. F. Gauss (1777-1855) with an improved least-squares method developed the basis of the regression analyses (Draper & Smith, 1998; Rouand, 2013). Regression analyses were made in 1869 by English statistician F. Galton (1822-1911), who in his anthropometric studies of human growth observed the average height of children (Ramcharan, 2006), depending on the average height of their parents (Hazewinkel, 2001; Rossman, 2013).

Although there is a greater number of regression analyses, with some authors citing 7, 12, 15 or even 40 regression analyzes, it is usually calculated with linear regression analyses (Sen & Srivastava, 2011). By studying various practical examples, the author developed a procedure for selecting significant regression analyses with a wide application of 16 different regression analyses, according to the following characteristics: quantitative or qualitative, linear or nonlinear, single or multiple, significant or special (Tofallis, 2009; Pan, 2009).

While the earlier calculation of regression analyses was carried out manually for hours or by special electromechanical devices, today the results are quickly obtained after only a few minutes, using modern electronic computers. By considering various practical examples, the author also defined the Procedure for Calculating and Applying Regression Analyses, to be applied for optimal problem solving, using Minitab® and Companion by Minitab® and electronic computers (Popović, 2019; Yoo & Harman, 2012; Sen & Srivastava, 2011; Yoo, 2012; Fotheringham, 2002).

THE PROCEDURE OF SELECTION ANALYSES

The method of calculating and applying regression analyses facilitates differentiated regression analyzes. Regression analyses can be divided into: parametric and nonparametric, quantitative and qualitative, linear and nonlinear, single and multiple, and more significant and special. Parametric regression analyzes have a finite number of factors, while non-parameters have an infinite number of factors (Popović & Ivanović, 2011).

Quantitative (Continuous) regression analyzes have a continuous dependent variable, an infinite brochure ($-\infty \leq X < +\infty$) of a numerical value and with quantitative or qualitative factors. Qualitative (Discrete) regression analyzes have an interrupt dependent variable, with the final a number of values ($X: 0, 1, 2, \dots, n$) or categories (Categorical), with quantitative or qualitative factors (Freedman, 2005; Laguna & Marklund, 2013).

Qualitative changes can be Binary (two categories: yes/no), Nominal (three or more categories without natural order, e.g. blue, red, white) and Ordinal (three or more categories in the natural order, e.g. cold, lukewarm, hot). Linear regression analyzes have linear coefficients in the regression function, the linear model and the straight line regression line. Nonlinear regression analyzes have nonlinear coefficients in the regression function, a nonlinear model and a curvilinear regression line (Yoo & Harman, 2012). Simple regression analysis has only one factor while multiple regression analyzes have multiple factors. The polynomial regression analyzes have a regression function with a constant and with one member (Mono), two members (Binomial), three members (Trinomial) (Hazewinkel, 2001).

More significant regression analyzes can be easily solved using the least squares method, using computer programs and electronic computers. Special regression analyzes, such as Ridge, LASSO and ElasticNet regression analyzes, are used to solve problems with the phenomena of multicollinearity and non-homogeneity of the variance, using the regularization procedure (Draper & Smith, 1998).

Significant quantitative regression analyzes are: 1. linear simple (CL- LINEAR), 2. linear simple orthogonal (CL- ORTHOG), 3. linear simple weight (CL- WEIGHT), 4. linear multiple (CL- MULTIPLE), 5. linear multiple stepwise (CL- STEPWISE), 6. linear multiple Partial (CL- PARTIAL), 7. linear multiple stability (CL- STABILITY), 8. non-linear simple (CN- SIMPLE), 9. non-linear multiple (CN- MULTIPLE) and 10. non-linear multiple polynomial (CN- POLYNOMIAL). Significant qualitative regression analyzes are: 11. linear simple binary (DL- BINARY), 12. linear simple nominal (DL- NOMINAL), 13. linear multiple ordinal (DL- ORDINAL), 14. linear multiple Poisson (DL- POISSON), 15. non-linear simple

(DN-SIMPLE) and 16. non-linear multiple (DN-MULTIPLE). If the existence of various types of regression analyses is accepted, then a suitable analysis is necessary, since the existing instructions are quite complicated (Freedman, 2005).

The procedure for selecting a significant regression analyses according to Figure 1: a) the detection of problems and the dependent size (quantitative or qualitative), b) detecting independent factors (one or more of them, with linear or non-linear influence on the dependent size), c) if a quantitative dependent quantity and one independent factor with a linear influence are detected, the regression analyses of CL-LINEAR, CL-ORTHO or CL-WEIGHT is chosen, d) if a quantitative dependent quantity and several independent factors with a linear influence are detected, the regression analyses CL-MULTIPLE, CL-STEPWISE, CL-PARTIAL or CL-STABILITY is selected, e) if a quantitative size and one factor with a non-linear effect is detected

then the CN-MULTIPLE regression analyses is selected, f) if quantitative size and several factors are observed with non-linear influence, then the regression analyses of CN-MULTIPLE or CN-POLYNOMIAL, g) if a qualitative quantity and one factor with a linear influence is detected then the DL-BINARY regression analysis is selected, h) if a qualitative quantity and several factors with a linear influence are observed, regression analyses DL-ORDINAL, DL-NOMINAL or DL-POISSON is selected, i) if a qualitative quantity and one factor with a non-linear effect is detected, then the regression analysis of DN-SIMPLE, j) if a qualitative quantity and several factors are observed with non-linear influence, then the regression analysis of DN-MULTIPLE, k) if the regression analysis does not have a good adjustment to the entered data, it should be returned to step detecting independent factors and l) if the regression analyses have a good adjustment, the selection procedure is completed (Popović, 2019).

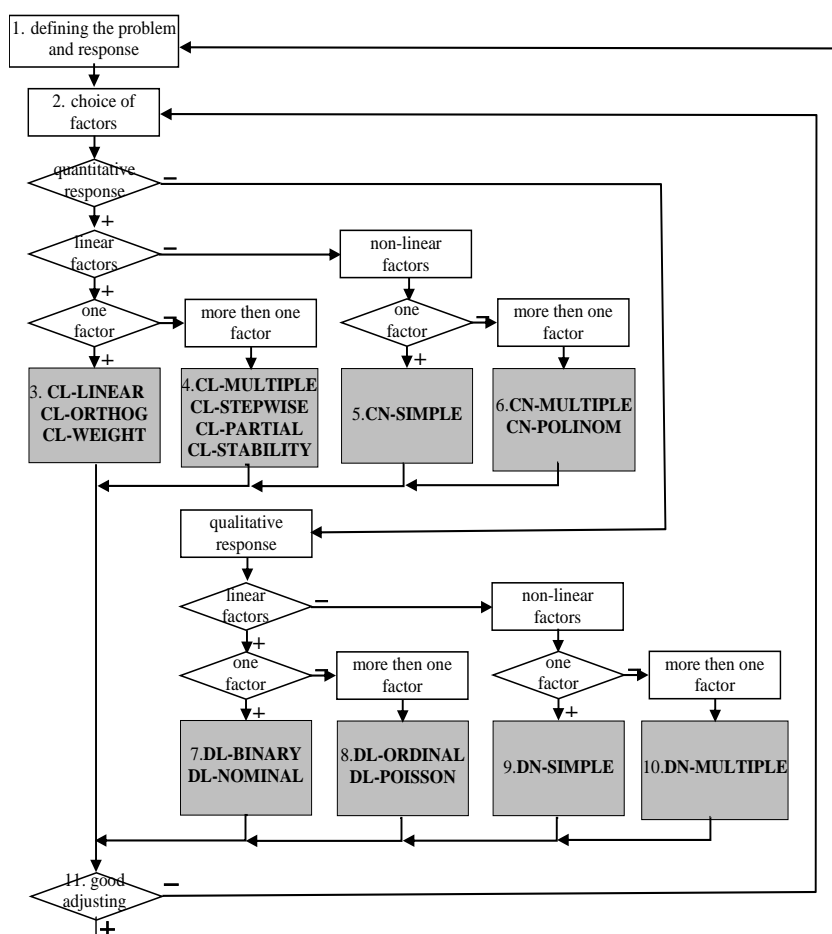


Figure 1: The procedure for selecting significant regression analysis

PROCEDURE FOR SOLVING ANALYSIS

If the necessary selection of a suitable regression analysis is made, then it is switched to the Procedure for calculating regression analyzes, using the Minitab® and Companion by Minitab® computer programs. Procedures for solving regression analysis involve the application of the following required steps: 1. Defining the problem (causes and consequences), 2. Defining the type of dependent size (quantitative or qualitative), 3. Defining the number of possible independent factors, 4. Selection of trial regression analysis (single or multiple), 5. Determination of the number of independent factors, 6. Taking a preferred sample size, 7. Calculation of basic statistical parameters, 8. Check deviation of calculated sample data, 9. Checking the choice of independent factors with a possible return to step 4, 10. The selection of the best selected independent factors, 11. Identifying the value of the dependent size (quantitative or qualitative), 12. Identifying the influence of independent factors (linear or non-linear), 13. Calculating the partial optimum according to the sample taken, 14. Checking the data adjustment with the possible return to step 3, if bad adjustments are detected (insufficiently or excessively), 15. Predicting useful future results, 16. Calculating a possible complete Optimus: simulating a large number of data, optimizing existing parameters, or improving the problem-solving process, and 17. Making unambiguous conclusions about the conducted regression analysis (Popović, 2019).

EXAMPLE OF SOLVING THE PROBLEM

The advantages of the recommended "The procedure for selecting a significant regression analysis" between 16 different regression analyses and "The procedures for solving regression analyses" are best seen when considering an illustrative practical example (Popović, 2019).

Problem: Solve the problem of finding the best sun exposure positions in order to achieve a heat flow increase (HeatFlux) based on the insolation values (Insolation) in East (East), South (South), and North (North) during a certain time (Time). with the lower limit of the tolerance of the required heat flow $LSL = 200 \text{ W/m}^2$.

In step a) detection of the problem and the dependent size determines the type of quantitative response. In step b) detection of independent

factors has determined several linear independent factors. In step c) The CL-MULTIPLE, CL-STEPWISE, CL-PARTIAL, and CL-STABILITY are detection for the type of regression analysis, due to the quantitative response and multiple linear independent factors. In step d) regression analyses detection by the determined group of analyzes was determined by regression analysis CL-MULTIPLE.

In step 1 Defining the causes and consequences of the problem, the following causes were identified: Insolation, East, South, North and Time and consequence: total heat flux Heatflux [W/m^2].

In step 2 Defining the type of response defined the type of quantitative response: Heatflux, y .

In Step 3 Defining the number of possible independent factors has determined the number of 5 possible independent factors: Insolation, East, South, North, and Time.

In step 4. Determination of the trial regression analysis according to the number of possible independent factors is determined 4. Quantitative linear multiple regression analysis (CL-MULTIPLE).

In Step 5. Determining the number of independent factors, applying the recommendations (Good and Hardin, 2009) with a ratio ($N = m^F$), where: N estimated size of the basic set of data, n the preferred sample size, and F the maximum number of independent factors, $N = 10^7$ and size $n = 30$, the maximum number of independent factors is obtained: $F = \log N / \log n = \log 10^7 / \log 30 = 7 / 1,477 = 4,739 \approx 5$.

In step 6. Taking the desired sample size, a mean sample of $n = 29$ measured values was taken, which is not shown here.

In step 7. The calculation of the basic statistical parameters obtained the following sample parameters: Insolation (754.5; 79.8), East (35.097; 1.417), South (35.529; 1.853), North (16.702; 0.908), Time (13.227; 1.947) and HeatFlux (249.64; 22.90).

In step 8, A deviation check of the calculated sample data was considered in the probability plot in Figure 2, which recorded the following: most point-to-point data is located in the middle line environment and within the drawn line curve with

a selected confidence interval of 95%, the percentage of the basic set in all samples are 95%, and the only factor North and response Heatflux do

not have a theoretical Normal distribution, due to the criteria for testing the hypothesis $P \leq (\alpha = 0.05) \rightarrow H_1$.

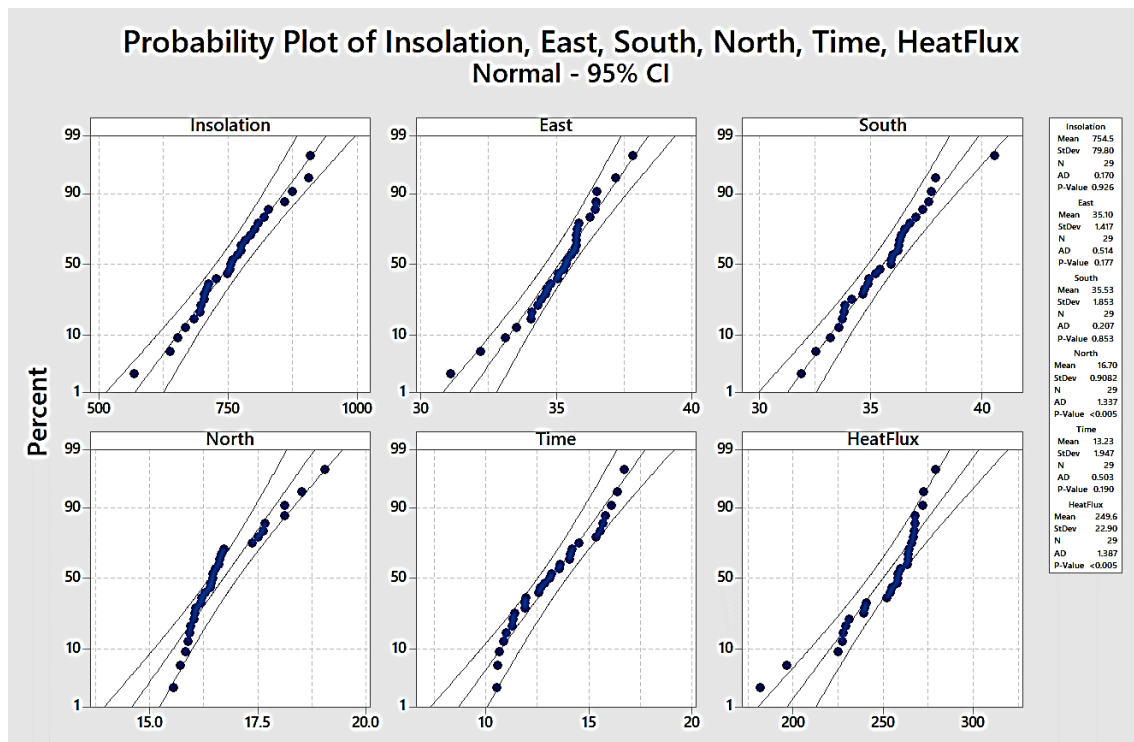


Figure 2: Probability Plot of factors

In step 9. To check the choice of independent factors, with the possible return to step 4, it is noted that: higher values of the factors in the sample give greater values of the response, that the factors have relatively low values of the deviation standards and are more likely to have higher factor values on a probability diagram, so there's no need to go back to step 4.

In step 10. The selection of the best selected independent factors, the table in Figure 3, was obtained with the following results:

- variant 5 with five factors is the best with the highest percentage of determination $R-sq=89.9\%$; $R-sq\ adj=87.7\%$, the coefficient $Cp=6.0$ is close to the number of factors and the standard deviation is small $S=8,0390$ while the other variants are worse,
- variant 4 has fewer percentages $R-sq=89.1\%$, $R-sq\ adj=87.3\%$, $Cp=5.8$ is greater than the number of factors and standard deviation is higher $S=8.1698$, and
- variant 3 has at least a percentage $R-sq=87.4\%$, $R-sq\ adj=85.9\%$, $Cp=7.6$ is greater than the number of factors and the standard deviation is the highest $S=8.5978$.

Based on these results, variant 3 was selected with only 3 independent factors: East, South and North.

Best Subsets Regression: HeatFlux versus Insolation, East, ...

Response is HeatFlux

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	Selection
1	72.1	71.0	66.9	38.5	12.328	
1	39.4	37.1	26.3	112.7	18.154	x
2	85.9	84.8	81.4	9.1	8.9321	x x
2	82.0	80.6	74.2	17.8	10.076	x x
3	87.4	85.9	79.0	7.6	8.5978	x x x
3	86.5	84.9	81.4	9.7	8.9110	x x x
4	89.1	87.3	80.6	5.8	8.1698	x x x x
4	88.0	86.0	79.3	8.2	8.5550	x x x x
5	89.9	87.7	78.8	6.0	8.0390	x x x x x

Figure 3: Selection response of HeatFlux

In step 11. Identifying the value of the response, the quantitative values of the response of Heatflux were determined: 271.8; 264.0; 238.8; .. ; [W/m²].

In step 12. Identifying the influence of independent factors, the linear effect of the factor is determined

because higher values of the factors in the sample cause larger values of the response.

In step 13. Calculating the partial optimum according to the sample taken, the results in the table in Figure 4 are obtained which show:

- good data adjustment along the middle line with the criterion $P \leq (\alpha = 0.05) \rightarrow H_1$ because the coefficients affect the response,

- high percentages of coefficients that prove the correctness of the chosen model, good adjustment with the factor *VIF* close unit,
- define the regression analysis function:

$$\text{HeatFlux} = 389.2 + 2.12 \text{ East} + 5.318 \text{ South} - 24.13 \text{ North} \quad (2)$$

Regression Analysis: HeatFlux versus East, South, North

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	12833.9	4278.0	57.87	0.000
East	1	226.3	226.3	3.06	0.092
South	1	2255.1	2255.1	30.51	0.000
North	1	12330.6	12330.6	166.80	0.000
Error	25	1848.1	73.9		
Total	28	14681.9			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8.59782	87.41%	85.90%	78.96%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	389.2	66.1	5.89	0.000	
East	2.12	1.21	1.75	0.092	1.12
South	5.318	0.963	5.52	0.000	1.21
North	-24.13	1.87	-12.92	0.000	1.09

Regression Equation

$$\text{HeatFlux} = 389.2 + 2.12 \text{ East} + 5.318 \text{ South} - 24.13 \text{ North}$$

Fits and Diagnostics for Unusual Observations

Obs	HeatFlux	Fit	Resid	Std Resid
4	230.70	210.20	20.50	2.94 R
22	254.50	237.16	17.34	2.32 R

R Large residual

Figure 4: Analysis of variance results

The obtained graphical results in Figure 5 show:

- probability plot of residuals with the adjustment to Normal distribution,
- versus fits residuals of the adjustment without the harmful phenomenon of non-homogeneity of the variance,
- residuals histogram with extreme values at the ends and
- residuals versus the order of data difference with great differences between 4th and 22nd data.

In step 14. Checking the data settings with a possible return to step 3 shows:

- progressive factor impact (South),
- degressive significant factor impact (North) with the criteria $P \leq (\alpha = 0.05) \rightarrow H_1$ and
- the influence of a low factor (East) with the criterion $P > (\alpha = 0.05) \rightarrow H_0$.

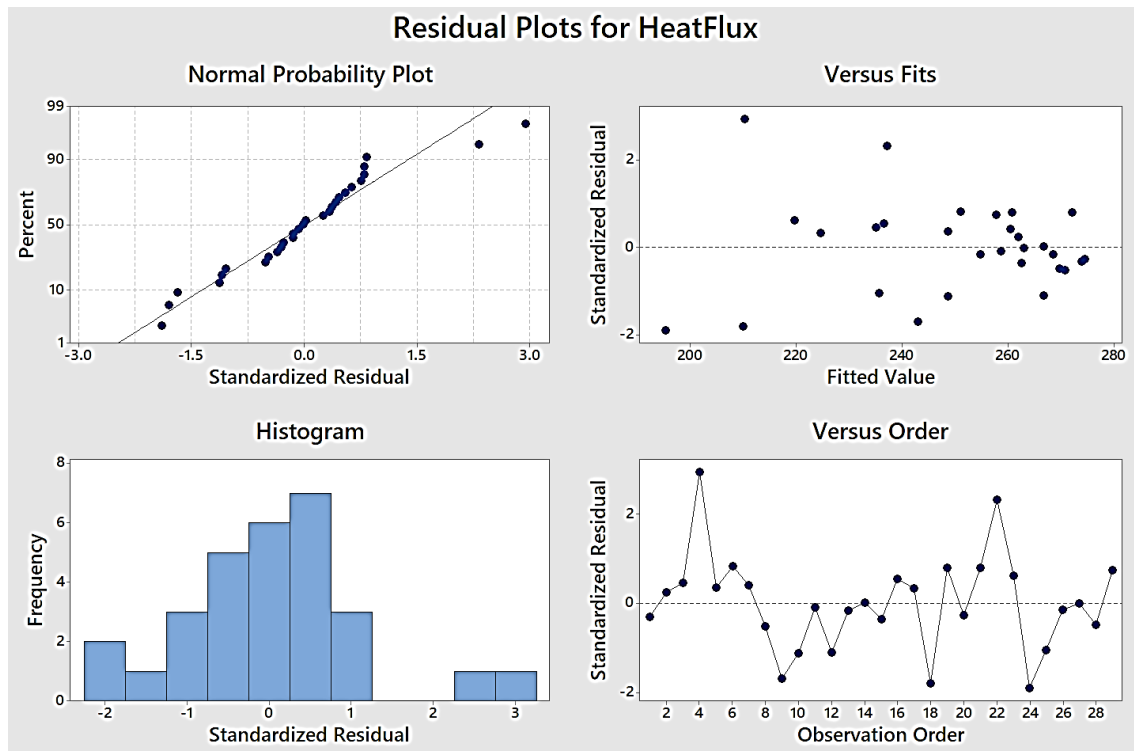


Figure 5: Residual Plots for HeatFlux

In step 15. The prediction of useful following responses gave the following prediction response result in Figure 6:

- prediction value: $HeatFlux = 206,783 \text{ W/m}^2$ with $SE \text{ Fit error} = 4,14474$,
- confidence interval obtained with a probability of 0.95 ($CI = 198.247 \div 215.320$) and
- predictive response interval depending on the future with a probability of 0.95 ($PI = 187.126 \div 226.441$).

In addition to the prediction of the following results, analyzes can be made here: interaction of regression analysis, contour diagram, surface diagram, border contours and optimization of responses.

Prediction for HeatFlux

Regression Equation

$$HeatFlux = 389.2 + 2.12 \text{ East} + 5.318 \text{ South} - 24.13 \text{ North}$$

Variable	Setting
East	36
South	33
North	18

Fit	SE Fit	95% CI	95% PI
206.783	4.14474	(198.247, 215.320)	(187.126, 226.441)

Figure 6: Predicting results for HeatFlux

In step 15a. The interaction diagrams of Figure 7 are obtained:

- the independent East factor has the least steep lines and the weakest impact on the response,
- the independent South factor has a steeper line and shows a greater impact on the response,
- the independent factor North has the most strident line and shows the greatest impact on the response of HeatFlux

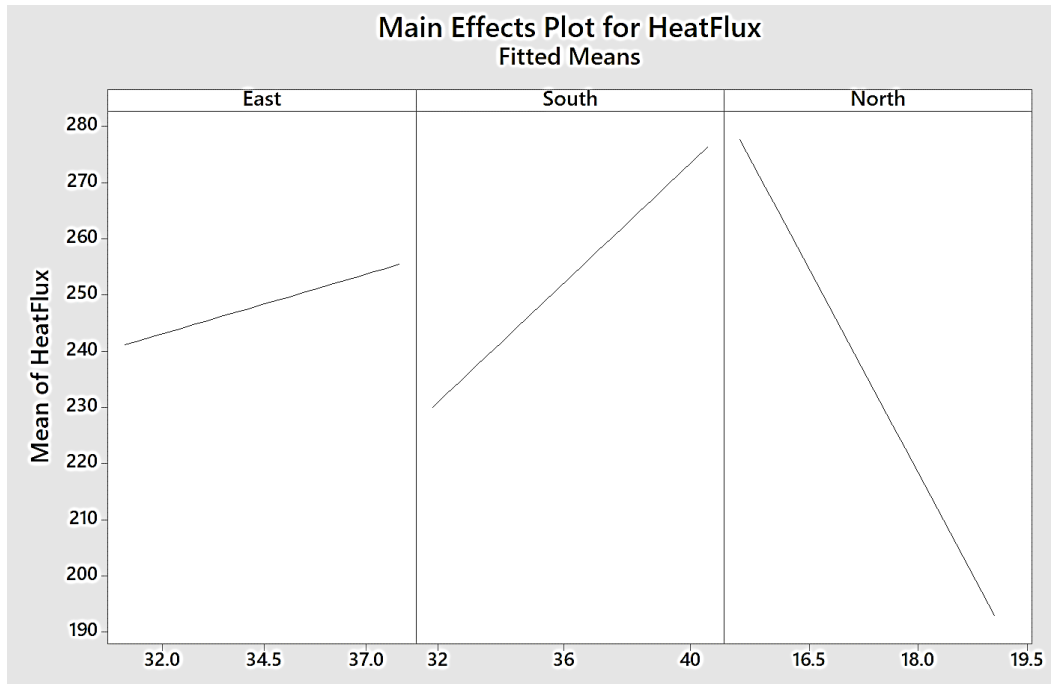


Figure 7: Main Effect Plot for HeatFlux

In step 15b. The contour diagram in Figure 8. was obtained:

- the North and South values are shown and the third East factor is not displayed, but its constant value is 35.096 for calculating the response value,
- the highest value (> 300) of the heat flow in the upper left corner of the diagram, which

corresponds to the high values of North (40) and South (15,5) and

- the lowest values (<180) of the heat flow are found in the lower right corner of the diagram, corresponding to the low values of North (19) and South (32).

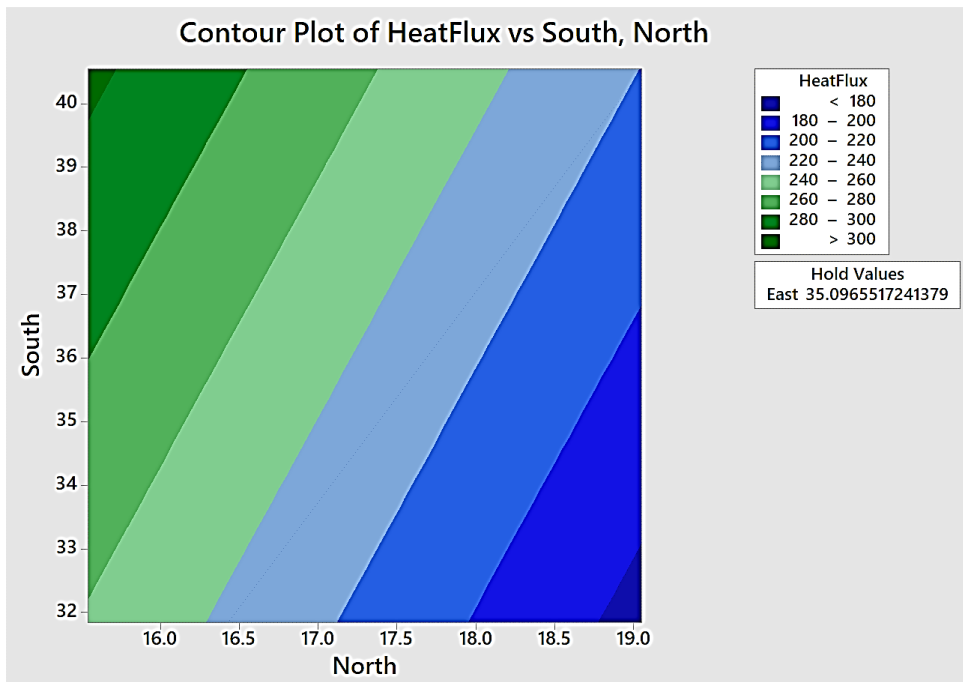


Figure 8: Contour Plot of HeatFlux vs South and North

In step 15c. The surface diagram in Figure 9. was obtained with the North and South factors with a constant value of the factor East= 35,096:

- the largest heat fluxes (about 300) are at the position of the factors North = 40 and South = 15.5 in the upper and the left part of the diagram,

- the lowest heat fluxes (about 50) are found for the North (15) and South (32) factors in the lower and right part of the diagram and
- the independent factor North has the longest designs that indicate the greatest impact on the HeatFlux response.

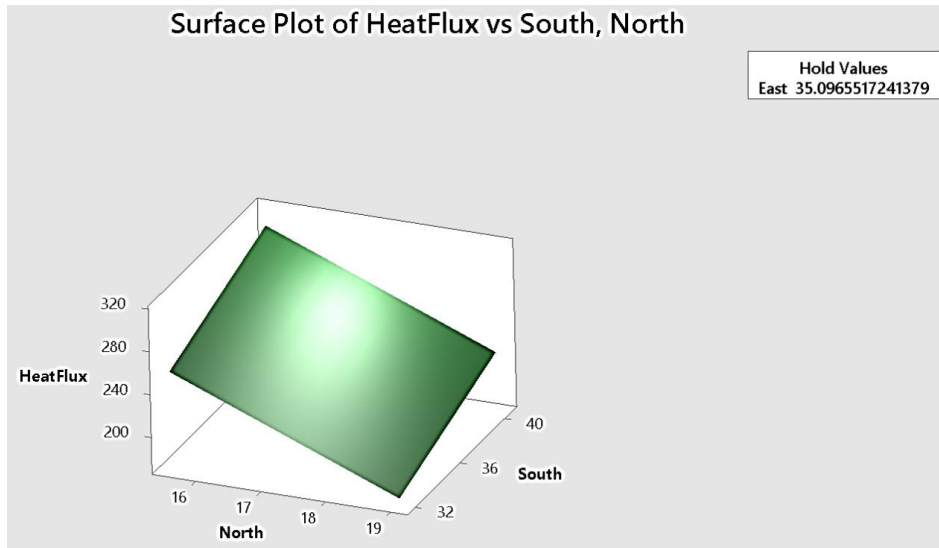


Figure 9: Surface Plot of HeatFlux vs South and North

In step 15d. The boundary contours of Insolation and HeatFlux values in Figure 10. were obtained, the values of the best positions of the heat flow with the East= 31.08 constant have the factors North and South:

- the maximum values of the North and South factors show full lines and
- the minimum values of the North and South factors are marked with broken lines.

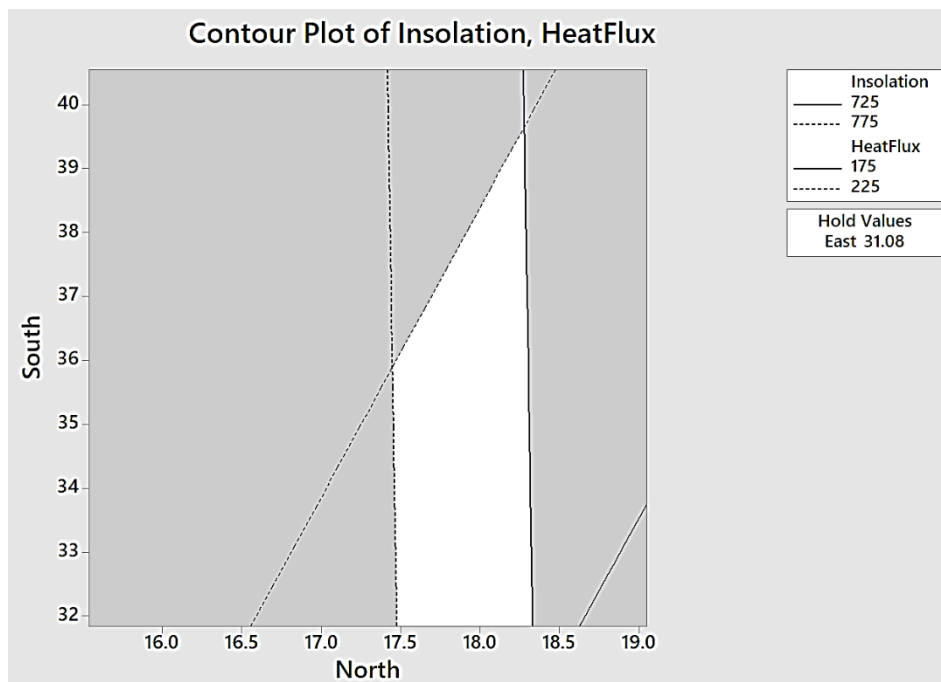


Figure 10: Contour Plot of Insolation, HeatFlux

In step 15e. Optimization of responses for Insolation and HeatFlux gives tabular and graphical results. The tabular results in Figure 11. are as follows:

- confidence interval (CI= 652.7 ÷ 847.3) with a probability of 0.95 and a prediction interval (PI= 591.3÷908.7) for Insolation and
- confidence interval (CI= 186.25 ÷ 213.75) with a probability of 0.95 and a prediction interval (PI= 177.58 ÷ 222.42) for HeatFlux

Response Optimization: Insolation, HeatFlux

Parameters

Response	Goal	Lower	Target	Upper	Weight	Importance
Insolation	Target	568.55	750	909.45	1	1
HeatFlux	Target	181.50	200	278.70	1	1

Solution

Solution	East	South	North	Insolation Fit	HeatFlux Fit	Composite Desirability
1	31.9438	31.84	17.6683	750	200	1

Multiple Response Prediction

Variable	Setting
East	31.9438
South	31.84
North	17.6683

Response	Fit	SE Fit	95% CI	95% PI
Insolation	750.0	47.3	(652.7, 847.3)	(591.3, 908.7)
HeatFlux	200.00	6.68	(186.25, 213.75)	(177.58, 222.42)

Figure 11: Response Optimiyation: Insolation and HeatFlux

The graphical results in Figure 12. are as follows:
- the optimum maximum values of solar radiation are in the East (37,820), South (40,550) and North (19,050) and

- the optimum minimum values for sunlight are in the East (31,080), South (31,840) and North (15,540) factors.

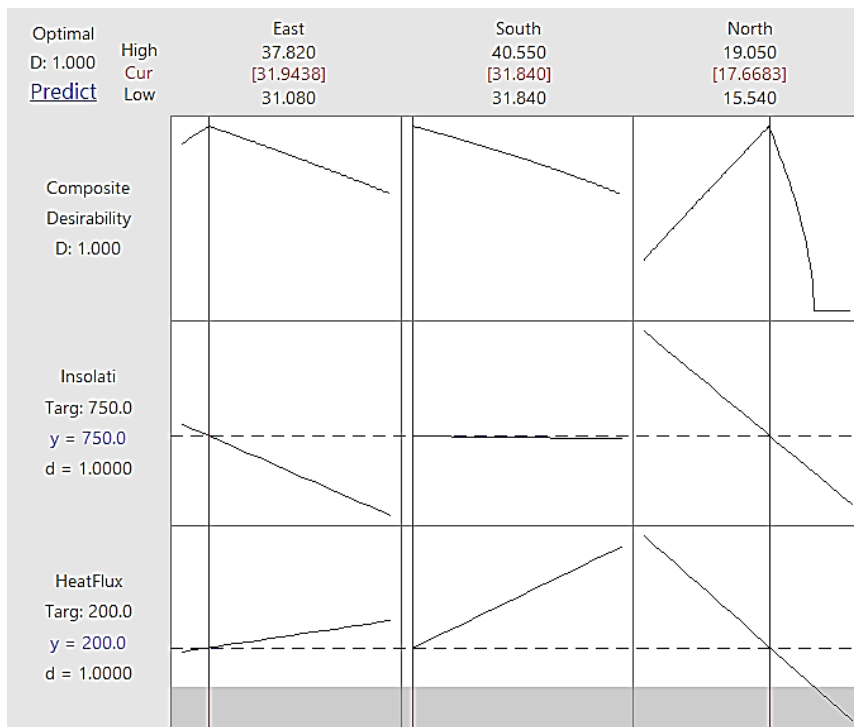


Figure 12: Optimal and Predict results

In step 16, Calculating the possible complete optimum here only allows the simulation of sample data ($n = 29$) with generated data obtained from a large sample ($n = 50\ 000$) using Monte Carlo simulation according to Figure 13. In Figure 14, the following simulation results were obtained:

- the accuracy index $PPI = 0.6791$, which is slightly lower than the exact process $PPI = 1.33$,

- below the lower limit of tolerance $LSL = 200$, there is only 2.04% of the heat flow and
- a large sample (50 000) gave a more realistic value of an average of 249,561 and a standard deviation of 24,107 heat flows

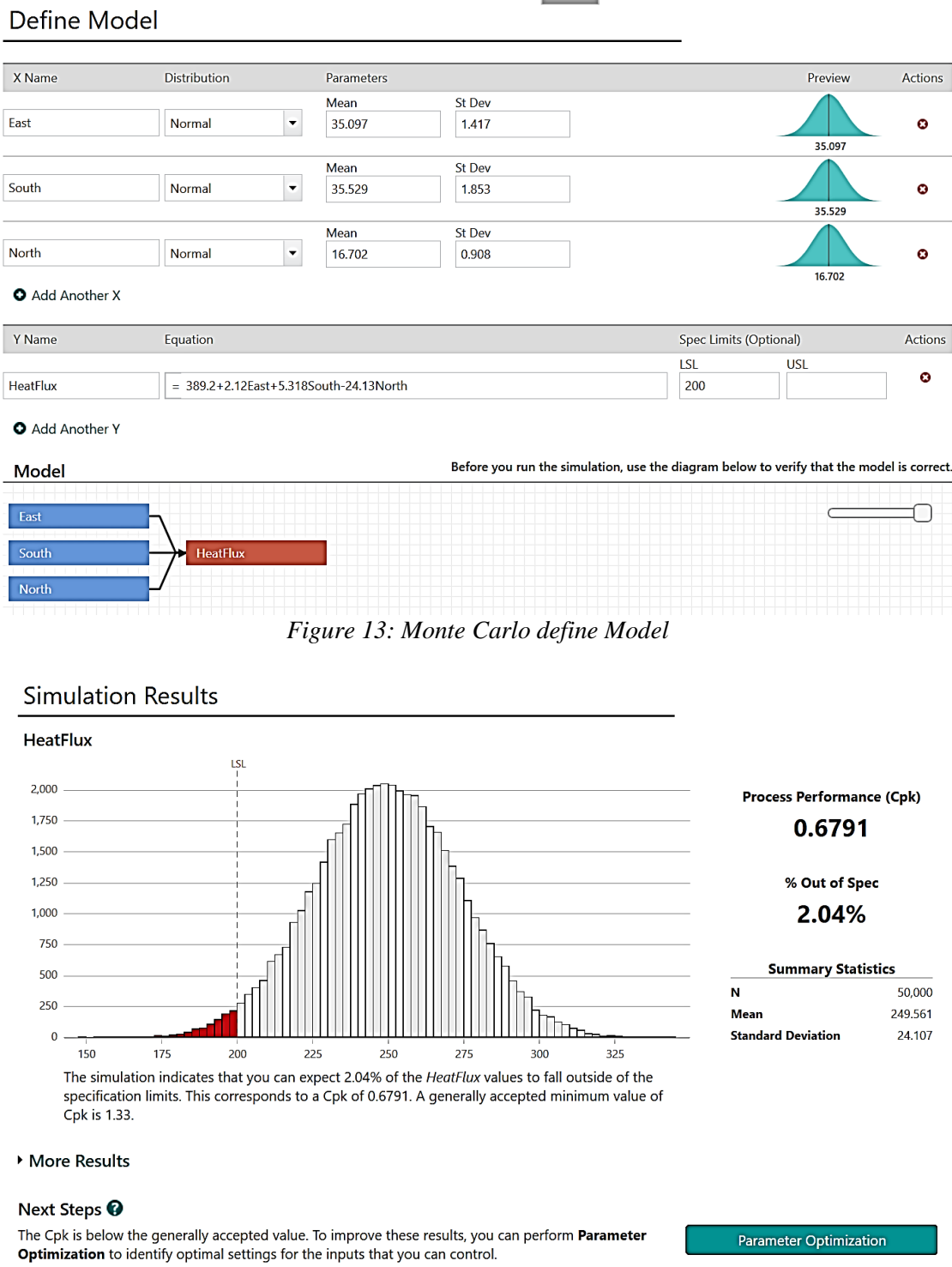


Figure 13: Monte Carlo define Model

Figure 14: Simulation results

In step 17. Making unequivocal conclusions about the conducted regression analysis, in order to solve the problems of the best insolation positions, based on the measured values of Insolation, East, South, and North, during Time encompasses:

- calculate the required regression dependency function: $HeatFlux [W/m^2] = 389.2 + 2.12 East + 5.318 South - 13 North$, and
- below the lower limit of tolerance (LSL= 200) there is only a little 2.04% of the flows and
- prediction $HeatFlux = 206,783 [W/m^2]$ with an increase in the resulting magnitude ($CI = 198,247 \div 215,320$) and the prediction interval of the future dependent size ($PI = 187,126 \div 226,441$) with a probability of 0.95.

CONCLUSIONS

The author pointed out that in everyday life and business as well there are often certain "problems" that hinder the achievement of the desired effectiveness in efficiency. These are various possible and impossible questions, tasks, and riddles that are being set up and attempt to solve in different ways. However, assuming that half of all these problems can be solved and that half of these problems can be mathematically defined, there is still a good chance of successful resolution.

In mathematics and statistics, problem-solving is used for regression analysis, which is now successfully solved by applying in different computer programs. The author suggests the application of special procedures "Procedure for selecting a significant regression analysis" between 16 different regression analyzes and "Procedure for resolving regression analysis" with an illustrative practical example of solving the problem of the optimal position of solar exposure to the solar collector.

REFERENCES

- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis* (3rd ed.): Wiley.
- Fotheringham, A., S., Brunson, Ch., & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. England, Chichester: John Wiley.
- Fox, J. (1997). *Applied Regression Analysis, Linear Models and Related Methods*: Sage.
- Freedman, D. A., & Statistical Models (2005). *Theory and Practice*: Cambridge University Press.
- Good, P. I., & Hardin, J. W. (2009). *Common Errors in Statistics (And How to Avoid Them)* (3rd ed.). New Jersey, Hoboken: Wiley.
- Hazewinkel, M. (2001). *Regression analysis, Encyclopedia of Mathematics*: Springer.
- Kirchmer, M. (2011). *High Performance through Process Excellence*. Berlin: Springer-Verlag.
- Laguna, M., & Marklund, J. (2013). *Business Process Modeling, Simulation and Design*: Taylor & Francis Group.
- Pan, L. (2009, August). Human age estimation by metric learning for regression problems. *In International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* (pp. 455-465). Springer, Berlin, Heidelberg.
- Popović B. (2017). *Simulacija poslovnih procesa – Simulation of business processes*. Beograd: Akademska misao.
- Popović B. (2019). *Optimalno rešavanje problema – The Optimal Solution to the Problems*. Beograd: Akademska misao.
- Popović B., & Ivanović G. (2011). *Sistem šest sigma u projektovanju rezultata procesa – Design for Six sigma*. Beograd: Mašinski fakultet.
- Ramcharan, R. (2006). Regressions: Why Are Economists Obsessed with Them? *Finance & Development*, 43.
- Rossman, A.J. (2013). *Workshop Statistics: Discovery with Data Minitab Companion*: Wiley.
- Rouaud, M. (2013). Probability, Statistics and Estimation: Propagation of Uncertainties in Experimental Measurement Retrieved from <http://www.incertitudes.fr/book.pdf>
- Sen, A., & Srivastava, M. (2011). *Regression Analysis – Theory, Methods, and Applications*. Berlin: Springer-Verlag.
- Tofallis, C. (2009). Least Squares Percentage Regression. *Journal of Modern Applied Statistical Methods*, 7(2), 526-534. doi:10.2139/ssm.1406472
- Yoo, S., & Harman, M. (2007). Regression Testing Minimisation, Selection and Prioritisation : A Survey. *Software testing, verification and reliability*, 00, 1-7. Retrieved from <http://www0.cs.ucl.ac.uk/staff/M.Harman/stvr-shin-survey.pdf> doi:10.1002/000

OPTIMALNO REŠAVANJE PROBLEMA PRIMENOM REGRESIONE ANALIZE

Problemi su različita moguća i nemoguća pitanja, zadaci i zagonetke koji su postavljeni i pokušani da se reše primenom različitih načina. Ipak, ako se pretpostavi da polovina svih problema nije rešiva i da se preostala jedna četvrtina ne može definisati, postoje dobre šanse za uspešno matematičko rešavanje druge četvrtine problema. U matematici i statistici, Regresione analize, koje se uspešno koriste danas korišćenjem različitih softvera, se koriste u rešavanju problema. Rešavanje problema korišćenjem regresionih analiza dozvoljava utvrđivanje odnosa između zavisnog odgovora i nezavisnih faktora kroz definisane regresione funkcije, procene veza, predviđajući rezultate, simulirajući podatke, optimizujući veličine i poboljavajući proces rešavanja. Iako postoje mnoge regresione analize, najčešće se koristi linearna regresiona analiza. Autor predlaže primenu Procedure za selekciju značajne regresione analize od 16 različitih regresionih analiza i Procedure za rešavanje regresionih analiza. Koristeći ovaj pristup, prikazano je optimalno rešenje za pronalaženje optimalne pozicije solarnog kolektora pri iskorišćenju sunčevog zračenja.

Ključne reči: Regresiona analiza, Optimalno rešavanje problema.