Exploring Regression Analysis in Predictive Modelling with Python tool

Deepak Kumar, Achin Srivastav, Manoj Kumar Sain, Vikash Gautam

Department of Mechanical Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur-302017 (India)

Email: **deepak.kumar@skit.ac.in, achin.srivastav@skit.ac.in, manoj.sain@skit.ac.in, vikash.gautam@skit.ac.in**

Received 02.09.2024 **received in revised form** 22.11.2024**, accepted** 27.11.2024

DOI: 10.47904/IJSKIT.14.2.2024.52-55

Abstract: This study investigates the application of Regression in predictive modeling by examining the relationships between 3 independent-variables (x1, x2 and x3) and 'y' as a dependent-variable. It shows the application of Regression in the domain of Plastic injection moulding, Welding, electric discharge machining and Casting. Utilizing a dataset with 22 observations, the analysis demonstrates that x1, x2, and x3 significantly influence y, with the regression model achieving an R-squared value between 75% and 90%. The results highlight the positive correlations between each independent variable and y, suggesting their substantial predictive power. The study underscores the effectiveness of Multi Linear Regression (MLR) in understanding complex variable interactions and provides insights into the model's fit and reliability.

Keywords- : **Regression analysis, Plastic injection moulding, Welding, Electric discharge machining and Casting**.

1. INTRODUCTION

Regression analysis is a fundamental statistical tool used in various fields, especially in manufacturing and engineering. It is a useful tool for predictive modeling. According to Winn-Nuñez et al., researchers understand the relationships between dependent (x) and independent variables (y), making it possible to predict the response variable, optimize the processes, and improve the part quality [1].

In manufacturing processes such as injection molding, welding, and casting, regression models are invaluable. They help to identify critical parameters that influence product quality and performance. By analysing these parameters, users can make data-driven decisions to reduce defects, enhance efficiency, and ensure consistent product quality. MLR (Multiple linear regression), in particular, provides a more comprehensive analysis by considering multiple independent variables simultaneously, offering deeper insights into complex processes and interactions. This leads to better control and optimization, ultimately resulting in cost savings and improved product quality and reliability.

The application of regression and multiple linear regression (MLR) techniques has been extensively studied in various manufacturing processes, providing valuable insights and optimization strategies.

a) **Injection Molding**: Research has shown that regression models can effectively predict and control the quality of injection-molded parts by analyzing key variables like injection pressure, melt & mould-temperature, and the cooling time. Studies highlight how MLR helps in optimizing these variables to minimize defects and improve product consistency [2-3]. Kumar et al. used Taguchi DOE technique and Regression tool to determine the thermal shrinkage, warpage and impact strength of Polypropylene material. Thus, with the help of Regression and Genetic Algorithm various defects were minimized [4].
b) **Welding**: In welding, regression analysis has been used to correlate welding parameters with weld quality metrics such as tensile strength and hardness. Literature highlights the role of MLR in identifying the optimal combination of welding speed, voltage, & current to achieve higher weld quality. It was investigated that in high quality SAW (Submerged-Arc-Welding) process, the weld-quality is greatly impacted by weld parameters like voltage between the arcs, welding-current in Amps. along with the distance between the nozzle and the plate and fast or slow speed of SAW welding. During the experiments, the Bead-on-plate welds were made by SS (polished stainless steel) plates by automated-SAW- welding process. During the collection of the data, RSM (response surface methodology) approach have been used which is an effective DOE technique. Along with this effective approach of RSM, statistical technique of regression analysis was also performed for dependent and independent variables [5]. During the research it was investigated that weld variables define micro level geometry of weld-bead and controls the mechanical properties like impact strength , fatigue strength and hardness of the fabricated joint. Regression analysis is also useful in laser welding which optimize the various parameters like laser power, speed of weld, focal position, shielding-gas-flow-rate, and the pulse-duration. Laser-Power and speed of weld controls the heat input along with the depth of penetration.

c) **Electric Discharge Machining (EDM)**: It was investigated that in EDM process the regression techniques generally used to model the relationship between process-parameters like pulse duration, discharge-current, and MRR (material removal rate). MLR has been particularly useful in optimization of these dominating parameters to enhance the machining efficiency and surface-finish.

According to Ishfaq et al., EDM was preferred for the precise machining in the production industries [6]. It was observed that fine and filtered kerosene oil used in EDM as hydrocarbon based which makes toxic-fumes and contributes to harmful emissions in the surroundings. Therefore, they examined potentiality of five biodegradable-dielectrics and Nano-copper powder used for cutting-proficiency of these dielectrics fluids. With the help of more than 20 experiments, regression analysis and ANN optimization for the MRR, the surface-finish, and the specific-energy-consumption was determined and thus high surface finish was obtained with good surface-finish. Thus, regression technique was found very effective for the optimization of EDM process[7].

d) **Casting:** In the various studies it was exhibited that in casting process, accurate linear regression models have been used to accurately predict the various casting-defects and improve dimensional accuracy. Research underscores the significance of MLR in understanding the effects of mold temperature, pouring rate, and cooling-rate on the final fabricated product which is the outcome of better-control and reduction of the defects.

It was investigated that in casting, key parameters include the mold-temperature, rate of pouring and rate of cooling, mold-material, & the alloy composition. Mold temperature affects the phase of solidification and it defects formation, while the rate of pouring influences filling and turbulence in the mould. Rate of cooling decides the macro and micro-structure and the mechanical-properties, and mold material and alloy composition significantly impact quality, surface finish and performance of the fabricated cast product[8].

In the study of Patel et al., it was shown that relationships between the variables in squeeze-casting process was best fit associated with the nonlinear regression models of response surface models, having the influence of the process-variable on different responses. This association has shown with effective surface graphs. As per the surface graph found, the squeeze-pressure, the die-temp., and pouring-temp. were recognized as effective variables that affect thickness and next level-dendrite arm-spacing. The next level-dendrite arm-spacing and thickness has shown linear collaboration with the squeeze-pressure and the duration of pressure, but curvy-linear function with pouring & hot die-temperatures[9].

In another study it was shown that the image-regression-analysis along with CNN (Convolutional ANN-neural network) were performed to correlate micro-structural- high definition images with various mechanical properties of polished martensitic-steel. It was investigated that the proposed CNN model can link micro-structure along with a property without any manual extraction of quality features [10]. Regression analysis can be understood by the following example which operated in the Python language [11].

2. REGRESSION ANALYSIS

For the analysis, x1 variable was taken in the range of 4-22 , x2 variable was in range of 2-23 and x3 variable was in range of 3-24. As shown in Table 1, the values of 'y' variable can be determined by conducting each experiment which is represented by each row. The data was entered in the python software in spyder notebook.

Table 1: Response values

Exp. No	x1	x2	x3	y	y_pred.	Residual
1	6	2	3	4.1195	2.68091	1.438648
2	8	3	4	3.0944	3.457815	-0.36338
3	7	4	5	2.5387	3.792044	-1.25325
4	4	5	6	4.3094	3.831156	0.478269
5	5	6	7	4.3139	4.460502	-0.14656
6	6	7	8	3.8876	5.089848	-1.20216
7	7	8	9	3.1825	5.719194	-2.53668
8	11	9	10	7.3602	6.34854	1.011707
9	9	10	11	6.2070	6.977886	-0.77085
10	10	11	12	7.7017	7	0.7017
11	11	12	13	8.6480	8.236579	0.411444
12	13	13	14	11.784	8.865925	2.918304
13	14	14	15	8.9976	9.49	-0.49767
14	15	15	16	10.893	10.12462	0.769329
15	16	16	17	12.768	10.753	2.014217
16	17	17	18	11.930	11.38331	0.547481
17	18	18	19	12.828	12.01266	0.816208
18	19	19	20	12.049	12.642	-0.59297
19	20	20	21	11.894	13.27135	-1.37675
20	21	21	22	11.591	13.90069	-2.30958
21	22	22	23	15.558	14.53004	1.02875
22	22	23	24	14.680	15.159	-0.47908

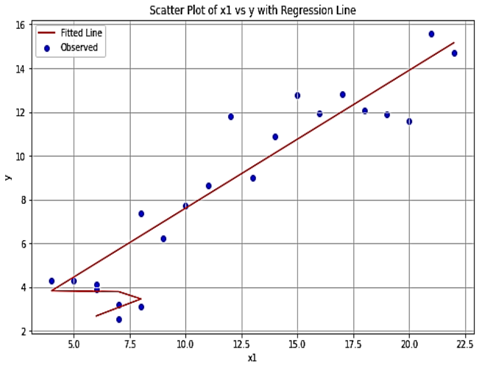


Figure 1: Response of x1 vs. y

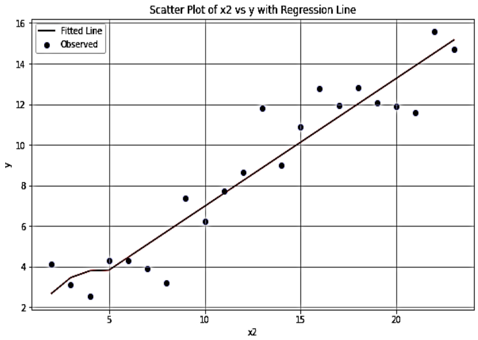


Figure 2: Response of x2 vs. y

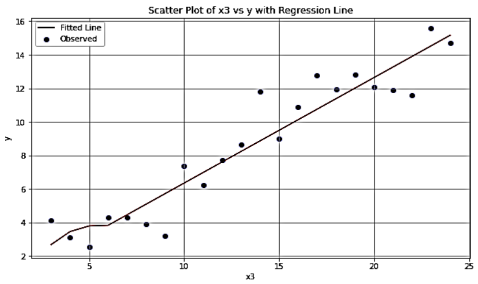


Figure 3: Response of x3 vs. y

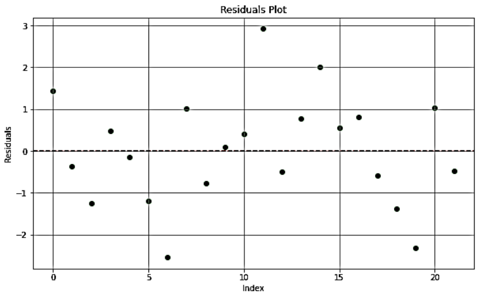
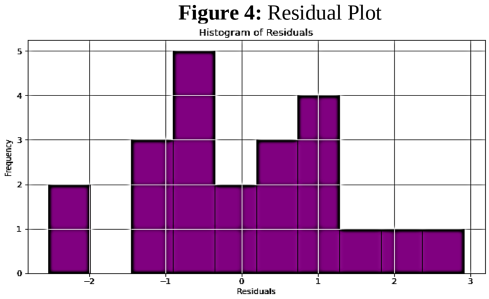


Figure 4: Residual Plot



Regression model: Response y = 1.04 -0.26*x_1 - 0.14*x_2 + 0.90*x_3 and R-squared: 0.90

Here's a concise analysis of relationship between the independent variables (x1, x2 and x3) and response variable 'y'.

1. **Variable x1 and Response y**: Fig 1 shows that the variable x1 has a positive correlation with y, indicating that as x1 increases, y increase as well. This suggests that x1 contributes positively to the response variable, making it a significant predictor.
2. **Variable x2 and Response y**: Similarly, x2 exhibits a positive relationship with y as shown in Fig 2. An increase in x2 generally leads to an increase in y, highlighting that x2 also plays a meaningful role in influencing y.
3. **Variable x3 and Response y** : The variable x3 is positively associated with y, reinforcing the same pattern observed with x1 and x2. It can be seen that as x3 rises, response y also increases, which indicates that x3 contributes positively to the dependent variable (y) as shown in Fig 3. Similarly, Fig 4 and Fig 5 shows the residual plot and residual histogram respectively.
4. **Multi-collinearity**: Although each independent variable shows a positive relationship with y, the extent of their individual contributions can be evaluated with their regression coefficients. It's important to assess potential multi-collinearity to ensure that the variables are not excessively intercorrelated.
5. **Fitness of the model**: The regression model's R-squared value falls within the user defined range (75-90%), showing that independent variables collectively describe a significant domain of the variability in y, and regression model provides a suitable fit for the data.

Thus, Python-Spyder, an integrated development environment (IDE), offers a powerful-platform for data- analytics. Its user-friendly interface facilitates the execution of complex code with features like an interactive console, variable explorer, and debugging tools. Spyder supports seamless integration with libraries such as pandas, numpy, and stats models, making it ideal for performing data manipulation and regression analysis.

CONCLUSION

In conclusion, it was observed that Regression analysis in the research area of Plastic injection moulding, Welding, Electric discharge machining and Casting is very useful and tuning of process parameters can be done with the help of Regression model. In this investigation, multiple linear regression (MLR) analysis conducted and showed a robust relationship between independent variables and response variable. The regression model achieves an R-squared value within the target range of 75-90%, indicating a strong explanatory power and a good-fit for the data. Each independent variable positively influences y, with increasing values of variables x1, x2, and x3 correlating with higher values of y. This suggests that these variables are significant predictors of the 'y' response. The study also highlights the importance of considering all variables collectively to understand their combined effect on y. Future research could explore the impact of additional factors for various manufacturing processes and address potential multi-collinearity issues to refine the model further. Thus, the study confirms the efficacy of Regression analysis in capturing the relationships among the variables and providing better insights.

REFERENCES

[1] E. T. Winn-Nuñez, M. Griffin, and L. Crawford, "A simple approach for local and global variable importance in nonlinear regression models," *Comput. Stat. Data Anal.*, (2024) , vol. 194, no. December 2023, p. 107914, doi: 10.1016/j.csda.2023.107914.

[2] Mollaei Ardestani, A., Azamirad, G., Shokrollahi, Y., Calaon, M., Hattel, J. H., Kulahci, M., Soltani, R., & Tosello, G. Application of Machine Learning for Prediction and Process Optimization—Case Study of Blush Defect in Plastic Injection Molding. Applied Sciences (Switzerland), (2023) 13(4). https://doi.org/10.3390/app13042617

[3] Maderthaner, J., Kugi, A., & Emmetmüller, W, Optimal control of the part mass for the injection molding process. Journal of Process Control, . (2023) 129,103027.https://doi.org/10.1016/j.jprocont.2023.10302

[4] D. Kumar, G. S. Dangayach, and P. N. Rao, "Enhancement of quality of polypropylene by optimisation of injection moulding parameters with genetic algorithm," *Int. J. Environ. Sustain. Dev.*, (2022), vol. 21, no. 1–2, pp. 206–217, doi: 10.1504/IJESD.2022.119389.

[5] A. Vedrtnam, G. Singh, and A. Kumar, "Optimizing submerged arc welding using response surface methodology, regression analysis, and genetic algorithm," *Def. Technol.*, (2018), vol. 14, no. 3, pp. 204–212, doi: 10.1016/j.dt.2018.01.008.

[6] K. Ishfaq, M. Sana, W. M. Ashraf, and V. Dua, "Sustainable EDM of Inconel 600 in Cu-mixed biodegradable dielectrics: Modelling and optimizing the process by artificial neural network for supporting net-zero from industry," *J. Clean. Prod.*, (2023), vol. 421, no. July, p. 138388, doi: 10.1016/j.jclepro.2023.138388.

[7] S, K., P, K., & G, V. Experimental Investigation of WE43(T6) magnesium metal matrix composites to enhance mechanical properties and EDM process parameters. International Journal of Electrochemical Science, (2024), 19(5), 100553. https://doi.org/10.1016/j.ijoes.2024.100553

[8] Mastrullo, R., Mauro, A. W., Pelella, F., & Viscito, L. , Process control and energy saving in the ladle stage of a metal casting process through physics-based and ANN-based modelling approaches. Applied Thermal Engineering (2024) 248(PA), 123135. https://doi.org/10.1016/j.applthermaleng.2024.123135

[9] M. Patel G.C., P. Krishna, and M. B. Parappagoudar, "Squeeze casting process modeling by a conventional statistical regression analysis approach," Appl. Math. Model., (2016), vol. 40, no. 15–16, pp. 6869–6888, doi: 10.1016/j.apm.2016.02.029.

[10] K. Sawai, T. Te Chen, F. Sun, T. Ogawa, and Y. Adachi, "Image regression analysis for linking the microstructure and property of steel," Results Mater., (2024), vol. 21, no. December p. 100526, doi: 10.1016/j.rinma.2023.100526.

[11] Kovač, N., Ratković, K., Farahani, H. and Watson, P., A practical applications guide to machine learning regression models in psychology with Python. Methods in Psychology, (2024) 11, p.100156.