

Role of Statistics in Engineering - A Review

Siddhant Banerjee¹, Utkrist Agrawal², Samisksha Agrawal³, Sonali Ankolikar⁴, Venkesh Agarwal⁵

¹School of Civil Engineering, Dr Vishwanath Karad MIT World Peace University, Pune-411038, Maharashtra, India

²School of Computer Science Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune-411038, Maharashtra, India

³School of Electronics and Communications Engineering, Dr Vishwanath Karad MIT World Peace University, Pune-411038, Maharashtra, India

⁴School of Electronics and Communications Engineering, Dr Vishwanath Karad MIT World Peace University, Pune-411038, Maharashtra, India

⁵School of Mechanical Engineering, Dr Vishwanath Karad MIT World Peace University, Pune-411038, Maharashtra, India

Abstract-Statistics is a very powerful tool that can be applied to different domains of study like engineering, management, finance, scientific research, banking, economics, mathematics, astronomy, banking, meteorology, industries and businesses. Statistics is a proven tool for decision making encompassing a wide range of sectors. It can handle enormous data very effectively, manipulate it for user requirements such that users can draw required conclusions for decision making. Most basic concepts of statistics such as mean, median, mode, standard deviation, regression as well as advanced analysis tools such as monte carlo analysis, design of experiments, six sigma, z-test, t-test, kruskal wallis variance analysis, chi-square test etc are being widely used across different domains to find first hand solutions to problems, decision making, optimisation, planning and scheduling and data analysis. Engineering also finds widespread applications of statistics in different domains such as production planning, quality control and management, process control, measurement system error analysis, robustness analysis, risk assessment etc. This paper is aimed at reviewing the role of statistics in engineering through case studies in different fields of engineering.

Keywords: Statistics; engineering applications; six sigma; standard deviation; models

I. INTRODUCTION

These statistical concepts and methods can be divided into two categories, which are parametric methods and non-parametric methods. As the name suggests, parametric methods are used when the parameters of the distribution are explicitly specified and non-parametric methods are used when no parameters of distribution are specified.

A. Mean and standard deviation

Mean and standard deviation are basic statistical tools used to analyse data. Mean in other words is the average of the available data and is also a measure of central tendency. On the other hand standard deviation is the deviation of each data term from mean. It provides information on the spread of the data and subsequently the variance. Both mean and SD can be calculated using excel and other softwares enabling the computation of huge data sets. An illustration of the importance of these two tools is being provided

through a project on wireless communication (Zhan and Goulart 2009). In this project, Under four different test conditions six tests were carried out. The signal to noise ratio abbreviated as SNR, for the bandwidth of communication was determined using mean and SD. For eg. in Electronics domain in image processing, the standard deviation (Σ) provides a measure of the dispersion of image gray level intensities and can be understood as a measure of the power level of the alternating signal component acquired by the camera.

Test No	Condition 1	Condition 2	Condition 3	Condition 4
1	4	2	16.7	14.3
2	3.3	2.2	25	14.3
3	2.5	2.5	20	12.5
4	1.9	2.3	20	12.5
5	2.3	2	14.3	12.5
6	3.1	1.9	25	12.5
Mean	2.85	2.15	20.17	13.10
SD	0.76	0.23	4.32	0.93
SNR	3.73	9.52	4.67	14.09

From the above table, with the help of statistical analysis using mean and SD to determine SNR we can identify the best condition for wireless communication. In absence of these tools, condition 3 might be identified as the best condition however it is not true. Condition 4 is the most optimal condition wherein condition 3 is after 2. This was an example of how effective statistical models are for analysing available data. It corroborates our analysis and expected results.

B. Six Sigma

Six Sigma is an important concept given by American engineer Bill Smith. Six Sigma is a set of statistical tools used for process improvement within total quality

management (TQM). It simply means that having a process that will generate products or outputs which have less than 3.4 defective parts per million products. It has various advantages and is a proven methodology of process improvement used in industries. It reduces costs, improves the results and data integrity, and improves the overall product lifecycle. Six sigma reduces the variability by eliminating the root causes of defects and thus prevents the defects in output of a process.

C. Regression

Regression is a method in statistics to identify the relationship between the variables in the data. It primarily focuses on the relationship between the dependent and independent variables which are also called predictors. It aids in understanding the changes occurring in the value of dependent variables when any one of the independent variables is changed and therefore giving us a relation between the two concerned variables. With the value of the independent variable, an equation is formulated. It contains the independent variables along with some coefficients with the slope value. There are several different types of regression techniques. One such type is linear regression technique that is mainly used for prediction. The simple linear regression with one dependent variable and one independent variable can be represented by :

$$y = c + m * x \quad (1.1)$$

where ,

- y is the predicted dependent variable value ,
- c is constant,
- m is coefficient of the regression ,
- x is the value of the independent variable.

There are many other types of linear regression techniques: Simple linear regression, Multiple linear regression, Logistic regression, Ordinal regression, Multinomial regression and Discriminant analysis.

Usually[4], analog circuit design relies on mathematical models able to characterize given circuit performances (gain, phase margin, slew rate, offset voltage, etc) as function of the design variables +(voltages, current, transistor sizes), with the aim to get a starting design point close to the desired circuit performance. However, such models can be very hard to get, especially if the circuit is non-linear and when the simulation time for the circuit is very high.

The regression models[1] are used to get the mathematical models for design variables in the form of a function from a training data set. They can be used as predictors and hence have been used plenty of times in analog circuit designing for parasitic modeling , verification and reliability analysis (including optimisation) etc. All in all, in these examples, the use of regression models allows the individuals to save simulation time by modeling the performances and nonlinearities.

In a regression model, the input given is the training data set which is composed of n samples such that[1]-

$$\{x_i, y_i\}_{i=1, \dots, n}, \quad (1.2)$$

where x_i is a vector with the design variables values

The x_i can also be referred to as the design point and y_i is the circuit performance corresponding to the i-th design point. The result is a model that can predict a value y_i as a function of x_i in which the difference error is compared with the actual y_i . There also exists another set of data called the testing data which is similar to the training data in terms of structure but is used to evaluate the prediction error of a regression method. Following are some features of a regression method-

- Training error: difference error compared with the training data set. It is desired to be as low as possible ,ideally speaking 0.
- Build time: this is the time required for processing the training data set and building the regression model. It is desired to be less than the time required to get the training data.
- Testing error: difference error compared with the testing data set. It is desired to be as low as possible, ideally speaking it should be 0.
- testing time: time required for predicting new data by using a regression model and it depends on the number of data to predict.
- Variables space:it is the capability to handle different numbers of variables yet preserving a low error.

D. Analysis of Variance(ANOVA)

Analysis of variance abbreviated as ANOVA is a popular statistical method which compares various samples. The purpose of analysis of variance is to test for significant differences between class means. ANOVA is based on the following assumptions:

1. The observations are independent of one another.
2. The observations in each group come from a normal distribution.
3. The population variances in each group are the same (homoscedasticity)

This technique proves to be extremely useful in revealing crucial information particularly in interpreting experimental outcomes and in determining the influence of some factors on other processing parameters.

II. APPLICATIONS IN ENGINEERING

A. Mechanical Engineering Case study -

A study from the manufacturing subfield is selected to review the role of statistics in mechanical engineering. In this particular study [4], statistics played a very important role in translating a huge amount of data in order to identify parameters of high speed machining. Statistics was used to translate the data into key surface roughness parameters of a product that was machined during a die cast process as well as to identify which parameter was most influential. Die casting is a forming manufacturing process capable of mass production. In this process, metal is melted and the molten metal under high pressure is injected to the casting mold to produce metallic products. Metals such as aluminum,

copper, lead and zinc are popularly used in this process. This process can produce a wide variety of products such as pistons of engines, cylinder heads of engines, propellers, gears, bushings, valves and other automobile components. High-speed machining, also known as high-speed cutting, is a more modern concept where comparatively much higher cutting speeds can be attained. High speed machining results in shorter machining times due to high material removal rate (MRR), enhanced surface quality, and better dimensional accuracy.

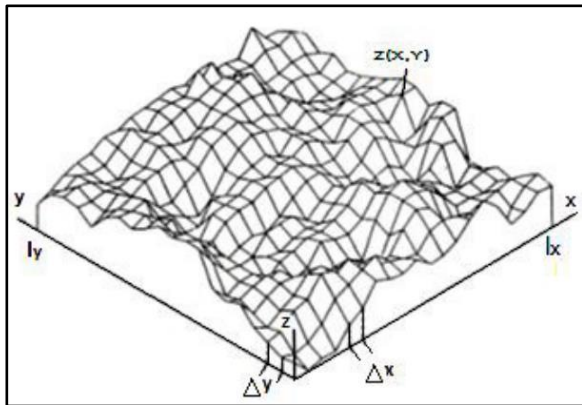


Fig.2.1.3D Surface roughness measurement, adapted from [4].

In this study [4], the surface roughness parameter is measured by measuring and then processing the texture data. Every roughness parameter is given by a numerical value which gives an estimate of surface roughness quantitatively. The study [4], was aimed at identifying the most crucial 3D surface roughness parameters and the most important technological parameters affecting roughness parameters. The HSM process was used on a milling machine having 16000 RPM maximum spindle speed, 26kW maximum power and 20m/min feed rate.

In order to identify the 3D surface roughness parameters, statistical analysis was used and a correlation matrix was prepared in which a set roughness parameters were inserted of n random variables in order to determine the most crucial roughness parameters. The statistical analysis resulted in identification of 5 key roughness parameters, which are

- a) Arithmetic mean surface height (Sa)
- b) Kurtosis of the surface (Sku)
- c) Height of the bearing area ratio (Stp)
- d) Texture aspect ratio (Str)
- e) Valley fluid retention index (Svi)

Next, in order to identify the most important technological factors affecting roughness parameters, again statistical analysis was used. The ANOVA multifactor analysis was used to achieve this aim. By replacing technological parameters with factors, and then conducting ANOVA multifactor analysis on pairs of these factors was the methodology used for analysis, The summary of results from this statistical analysis is summarized in Table 3.

Influence of technological parameters on the roughness parameters [4]

3D Parameter	Texture	Influenced by	ANOVA ratio
Sa	Strategy		0.002253
Sku	Material		0.01678
Stp	Material		0.03999
Str	Feed Rate		0.01621
Svi	Strategy		0.04633

Sa	Strategy	0.002253
Sku	Material	0.01678
Stp	Material	0.03999
Str	Feed Rate	0.01621
Svi	Strategy	0.04633

It is evident from the study, how important statistical analysis is in engineering. It facilitates analysis of huge amounts of data leading to significant findings, simplifies the process and yields much accurate results.

B. Applications of Statistics in Electronics and Communication -

1. In communication network[3]: In recent years there has been an immense increase in the need for efficient and powerful communication network systems. The gigantic boom in the network and connectivity industry has given rise to the use of optimisation in the sector to cater to the increasing global data traffic. Though the origins of statistical planning were initially related with the models and predictive supply and demand situations, the statistical models are being implemented to improve and enforce major communication networks to advance the high speed communication. In this area, we consider the number of nodes in the network and the respective nodal degrees to calculate the various distributions.

Applications in this field[3]:

- Number of links in specific ranges: It is seen that the number of links with lengths in a certain length range can be estimated using the expressions for link statistical model. Cumulative distribution function (CDF) as well as the area under the PDF of the distribution can be used to estimate the number of links within a specific range.
- Total no. of amplifiers: Since optical amplifiers are used to strengthen the signal and estimate the link dependent parameter. By using a statistical model, we can essentially create something that understands the importance of taking into account the distribution of link lengths and link partitions to reduce any errors and thereby

preventing miscalculation in the need of amplifiers.

- Types of Modulation Schemes at the Nodes :

The type of modulation and demodulation done at nodes depends on the link length and thus it makes more sense to use distance dependent modulation and demodulation to prevent excess efforts.

- Total Length of Fiber: The length of fiber used is directly related to the cost of fiber.

∴ Total fiber cost =
 Cost per unit length of fiber x average link length x number links

- Design and cost estimations

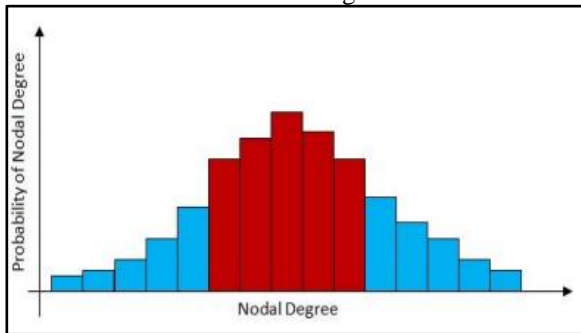


Fig.2.2[3] In these types of networks, the nodal degree values are very similar and follow Poisson's distribution. These nodal degrees take up a bell shape to prove and demonstrate how they follow up to Poisson's distribution property and characteristics. This is a property of "random" or "random like" topologies.

2. Statistics in Signals and Signal Processing: Concepts demonstrated above have been implemented to create Signal Processing using the domains concerned with networking and digital signal processing. This helps to generate visualization and improved data representation using various transforms. Laplace transforms, frequency domain transforms and more help in creating a graph equivalent.

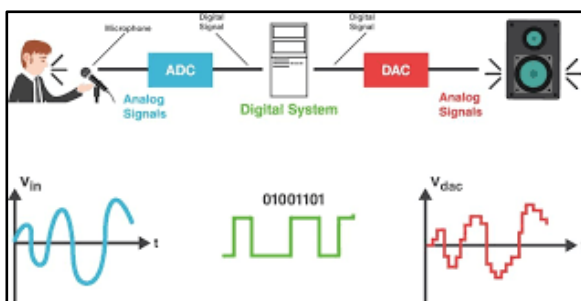


Fig.2.3. In the above image we can observe the phenomenon known as digital signal processing which is majorly used in conversion of various types of signals w.r.t their SNR and power levels.

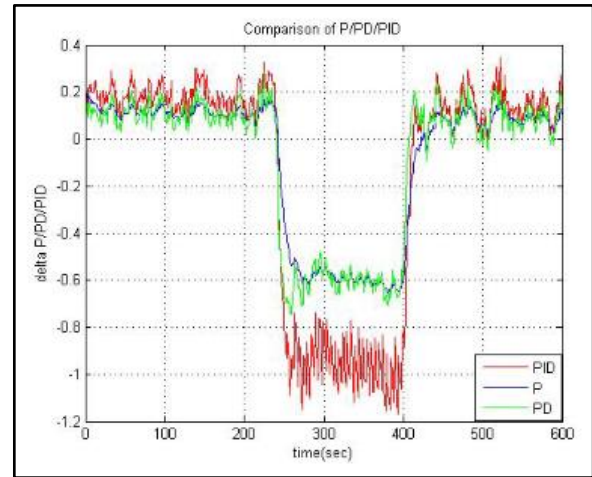


Fig.2.4. Comparison of P, PD, and PID objective functions where the functions are plotted in real time to demonstrate the different stabilities/ reliabilities of the channel selection algorithm in [2] Optimal channel selection in real time multimedia uplink transmissions in ambulances.

C) Applications of statistics in Civil Engineering

1. Structural Auditing with various methods

Online vibration-based harm recognition strategies are progressively well known for distinguishing harms during functional seasons of huge common and mechanical designs. Specifically, for complex public designs, for example, spans, where human wellbeing is vital, or for structures hard to get to and assess, similar to wind turbine sharp edges or seaward establishments. For given cases, among numerous different models, the vibration-put together harm recognition depends with respect to the ID of the harm-initiated deviations in the harm touchy amounts of the gathered reaction signals.

A regular practice is to utilize a modular methodology, which assumes that the harms are completely reflected by the vibrational attributes (regular frequencies, mode shapes or damping proportions) distinguished from the information and from that point thought about between the sound and present statuses. Be that as it may, hands on work questions the immediate utilization of the modular boundaries, contending that the modular information itself isn't sufficiently touchy to recognize the nearby blames, particularly when, by and by, the construction is invigorated by low-recurrence inputs. One detour to the modular system is to utilize the measurable techniques, where trademark harm delicate amounts are gotten straightforwardly from the information and assessed for harms in a speculation tests.

2. Mahalanobis distance-based damage detection

The square MD between the observations in the data vector \mathbf{x}_i and a reference, baseline model with the sample mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ is defined as:

$$MD_i = (\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \quad (2.1)$$

3. Numerical Simulations:

The mathematical tests are directed on a limited component (FE) model of a MB establishment - another idea for a help structure for seaward wind turbines. The MB structure consists of a roundabout steel shell shaping a skirt, which is introduced inside the seabed and shut with a roundabout plate that makes water/air proof conditions inside the supposed pail. The water/air proof element permits to introduce the establishment with attractions siphons, that is quiet and quick to accomplish. The shaft is associated with the establishment by steel profiles called networks, which move the functional burden to the skirt. The welded shaft-web association is inclined to high burdens and convey a critical weariness load, consequently it is considered as a potential harm area.

The primary reactions are reproduced by utilization of the FE model of the construction with a can breadth of 14 m and a 32 m long shaft. The translational and rotational limit conditions are compelled to zero on the skirt plates. In all, the FE model contains 8589 first-request shell components, 8414 hubs and, therefore, 50484 levels of opportunity (DOF). Yield speed increases are mimicked utilizing repetitive sound of difference taken haphazardly from an ordinarily dispersed vector in the middle, following up on the hubs on top of the shaft. A solitary age of the reaction is recorded for 250 s with an examining recurrence of 40 Hz in 5 hubs by bi-hub sensors, therefore yielding 10 speed increase channels. Altogether, the encompassing vibrations are reproduced for 45000 s, which brings about 180 informational collections; 50 sets from the solid state and 130 sets addressing 13 harmed situations. To challenge the exhibition of the harm identification strategies, 1% of a Gaussian background noise added to the reaction signals. The harms are reproduced as a dynamic thickness (t) decrease of the components in the shaft-web association, by, individually, 1%, 5%, 15%, 40% and 85%. Every component is a square of 100mm x 100mm. The harm test situation alongside a comparing informational index are portrayed in a table below.

Annotation	Description	Sets	Annotation	Description	Sets
A	Healthy state	50	H	t of 4 FE in ALL connections reduced by 15%	10
B	t of 4 FE in A reduced by 1%	10	I	t of 4 FE in A reduced by 40%	10
C	t of 4 FE in A reduced by 5%	10	J	t of 4 FE in A and B reduced by 40%	10
D	t of 4 FE in A and B reduced by 5%	10	K	t of 4 FE in ALL connections reduced by 40%	10
E	t of 4 FE in ALL connections reduced by 5%	10	L	t of 4 FE in A reduced by 85%	10
F	t of 4 FE in A reduced by 15%	10	M	t of 4 FE in A and B reduced by 85%	10
G	t of 4 FE in A and B reduced by 15%	10	N	t of 4 FE in connections reduced by 85%	10

Table.2.4: Damage scenarios during the simulations on the MB model.

3.1 Damage Detection Outcomes

The reference state is made utilizing the initial 30 informational indexes from the sound state. All harm identification tests are led with the boundary setting $q = p + 1 = 5$. The harm pointers for the mathematical experiments alongside the combination of subspace-based and MD-based strategies are delineated in the illustrations below.

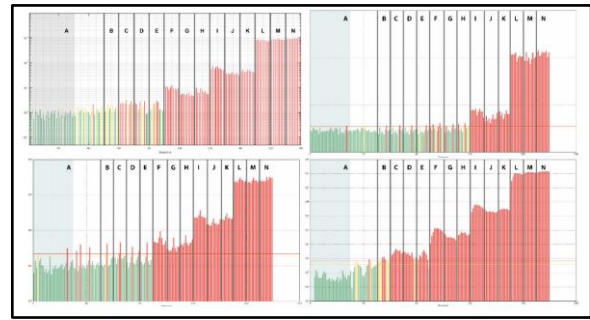


Fig.2.5. MD-based damage indicators (upper left). Classic subspace-based damage indicators (upper right). Robust subspace-based damage indicators (lower left). Fusion of the subspace-based and MD-based damage indicators in the Hotelling control chart (lower right)

Every one of the three techniques recognize the cases with 40 % and 85 % decrease of thickness, while the 15 % decrease is as it were recognized by the MD-based and the vigorous subspace-based strategies. The principal stage with 5 % decrease is identified by the MD-based calculation, nonetheless, alongside a few bogus alerts set off in the sound state. Just the combination of all the discovery techniques in the Hotelling control graph are equipped for recognizing each harm situation and don't lay out harms in the solid information.

III. CONCLUSION

In this paper we successfully studied the applications and inevitably important role of statistics in the various fields of engineering. The branch of Mechanical Engineering applies it for intense high speed machining study. It also helps in gathering large amounts of data to analyze die casting parameters. In electronics and communication engineering we observe its many sides and applications in helping with network analysis to help networking companies to improve their network connectivity, customer experience w.r.t service availability, etc. Also we observe it helps in providing many optimized solutions in laying groundwork for the analog circuitry. When it comes to civil engineering, we see it provides impressive analysis of structures and their various parameters. All in all, as the world expands and new data is generated in vast amounts everyday, statistics helps in creating important decision systems to improve the work engineering field does. A huge amount of research goes into the development of solutions apart from those discussed in this paper to create a better world with help of engineering and its various branches.

IV. FUTURE SCOPE

Essentially when we consider the scope for this paper, we pondered the different case studies of different applications of statistical tools in the field of engineering. With more intensive in-depth research we may come across the coming implementations of these methods in much more complex real world examples like the recent scenario of increase in demand of electronic devices in the recent era of online learning. We can see electronics, civil and mechanical engineering have vast expanse and innumerable applications

of statistics. With every set of data and information generated during experiments in strength testing etc, comes the role of statistics in helping it make sense or comprehensive. We may be able to add to this survey paper the newer research on recent decision making technologies used in each technology. In the age of 5G we can also add the importance of improved algorithms in network planning and management since more than half the world's youth draws their education from online sources and depend on improved network and net conditions for connectivity and education. The recent paradigm shift in creating smarter material has also increased and enabled the research to increase testing thus creating more scope for us to study and delve deeper into the material and mechanical properties.

REFERENCES

- [1] Ortega, Antonio, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst. "Graph signal processing: Overview, challenges, and applications." *Proceedings of the IEEE* 106, no. 5 (2018): 808-828.
- [2] Goulart, Ana, Wei Zhan, and Robert Arnold. "Optimal channel selection for real-time uplink data transmissions in ambulances." *Journal of Mobile Multimedia* (2009): 271-286.
- [3] Shreesha, S., and Sudhir K. Routray. "Statistical modeling for communication networks." In *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, pp. 831-834. IEEE, 2016.
- [4] Guerra-Gomez, Ivick; McConaghy, Trent; Tlelo-Cuautle, E. (2015). [IEEE 2015 16th Latin-American Test Symposium (LATS) - Puerto Vallarta, Mexico (2015.3.25-2015.3.27)] 2015 16th Latin-American Test Symposium (LATS) - Study of regression methodologies on analog circuit design. , (), 1-6. doi:10.1109/latw.2015.7102504
- [5] Kikuchi, Kazuro. "Fundamentals of coherent optical fiber communications." *Journal of Lightwave Technology* 34, no. 1 (2015): 157-179.
- [6] S. K. Routray, G. Sahin, J. R. F. da Rocha and A. N. Pinto, "Statistical Analysis and Modeling of Shortest Path Lengths in Optical Transport Networks," in *Journal of Lightwave Technology*, vol. 33, no. 13, pp. 2791-2801, 1 July1, 2015, doi: 10.1109/JLT.2015.2413674.
- [7] A. Singh, G. P. Indira, K. Monica, R. Tiwari, A. Kala and S. K. Routray, "Emerging characteristics in wireless sensor networks: Statistical viewpoints," 2017 International Conference on IoT and Application (ICIOT), 2017, pp. 1-4, doi: 10.1109/ICIOTA.2017.8073618.
- [8] M. Zaliskyi, O. Solomentsev, R. Odarchenko, O. Shcherbyna, L. Tereshchenko and V. Gnatvuk, "Sequential Estimation of Reliability Parameters of Telecommunication and Radioelectronic Systems," 2021 IEEE 16th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), 2021, pp. 6-9, doi: 10.1109/CADSM52681.2021.9385251
- [9] M. Döhler, F. Hille, L. Mevel, & W. Rücker, Structural health monitoring with statistical methods during progressive damage test of S101, Bridge, Engineering Structures, 69 (2014) 183-193.
- [10] M. D. Ulriksen, D. Tcherniak, P. H. Kirkegaard, L. Damkilde, Operational modal analysis and wavelet transformation for damage identification in wind turbine blades, Structural Health Monitoring, 4 (2016) 381-388.
- [11] W. Weijijens, T. Verbelen, G. De Sitter, C. Devriendt, Foundation structural health monitoring of an offshore wind turbine- a full-scale case study, Structural Health Monitoring, 4 (2016) 389-402.
- [12] S. Doebbling, C. Farrar, M. Prime, A summary review of vibration-based damage identification methods, Shock Vib. Dig. 30 (1998) 91-105.
- [13] G.T. Houlshby, B.W. Byrne, Suction Caisson Foundations for Offshore Wind Turbines and Anemometer Masts, Wind Engineering 24(2000),pp. 249-255
- [14] E. Carden, and P. Fanning, Vibration based condition monitoring: A review, Struct. Health Monit. 3 (2004) 355-37
- [15] Zhan, Wei & Fink, Rainer & Fang, Alex. (2011). Application Of Statistics In Engineering Technology Programs.American Journal of Engineering (AJEE).I. 10.19030/ajee.v1i1.793.
- [16] Saleem, Iram & Aslam, Muhammad & Azam, Muhammad. (2013). The Use of Statistical Methods in Mechanical Engineering. Research Journal of Applied Sciences Engineering and Technology. 5. 2327-2331. 10.19026/rjaset.5.4660.
- [17] Logins, Andris & Torims, Toms. (2015). The Influence of High-speed Milling Strategies on 3D Surface Roughness Parameters. Procedia Engineering. 100. 10.1016/j.proeng.2015.01.491.
- [18] Osama Ahmed Marzouk, Ahmad Izzat Jamrah. Case Studies of Statistical Analysis in Engineering. International Journal of Statistical Distributions and Applications. Vol. 3, No. 3, 2017, pp. 32-37. doi: 10.11648/j.ijstd.20170303.12
- [19] V. Agarwal and S. Ankolikar, "Deployment of RFID sensors in supply chain management – a review," *Journal of Mechatronics and Artificial Intelligence in Engineering*, Jul. 2022, doi: 10.21595/jmai.2022.22565.