Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 40-50, 2, 2024
© Telif hakkı IJANSER'e aittir
Araştırma Makalesi

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 40-50, 2, 2024
Copyright © 2024 IJANSER
Research Article

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

Addressing Missing Data in Surveys and Implementing Imputation Methods with SPSS

Robert Kosova*, Adrian Naço[2], Shkelqim Hajrulla[3], Anna Maria Kosova[4]

[1*] *Department of Mathematics, University "Aleksandër Moisiu" Durrës. Albania*
[2] *Department of Mathematics. Politechnic University of Tirana. Albania*
[3] *Computer Engineering Department. Epoka University. Tirana. Albania.*
[4] *Department of Computer Science. University "Aleksandër Moisiu" Durrës. Albania*

Email of corresponding author: robertkosova@uamd.edu.al

(Received: 03 March 2024, Accepted: 08 March 2024)

(4th International Artificial Intelligence and Data Science Congress ICADA 2024, March 14-15, 2024)

ATIF/REFERENCE: Kosova, R., Naço, A., Hajrulla, S. & Kosova, A. M. (2024). Addressing Missing Data in Surveys and Implementing Imputation Methods with SPSS. *International Journal of Advanced Natural Sciences and Engineering Researches, 8(2), 40-50.*

Abstract **–** The presence of missing data in surveys or in other types of scientific research poses significant challenges for academic research, impacting the reliability, validity, and generalizability of study findings. Missing data can introduce bias into the analysis, leading to erroneous results and conclusions. Furthermore, missing data can compromise the statistical power of analyses, reducing the precision and accuracy of estimates. Consequently, researchers may encounter difficulties in drawing robust conclusions or identifying significant patterns within the data. Missing data can stem from various sources, including participant non-response, data entry errors, survey design flaws, and respondent unwillingness to disclose sensitive information. Many researchers struggle with how to handle missing data in their studies. They often use simple methods like deleting all the cases with missing data, partial deletion, or filling in missing values with a single number, such as the values of the variables mean, median, or mode. However, these methods can be misleading because they don't take into account the reasons why data might be missing and produce estimation errors, so other approaches are implemented to impute the missing values. However, each approach has its limitations and assumptions, which can influence the validity of results and introduce additional uncertainty into the analysis. This article analyzes the problem of missing data in social surveys, the reasons for missing data, the types of missing data, and also suggests several ways, deterministic and probabilistic, for data imputation.

Keywords – Missing Data, Imputation, Survey, SPSS, Deterministic, Probabilistic

I. INTRODUCTION

Surveys represent a useful tool that is used in many fields, including social, political, and educational, with unique objectives and methodologies. Surveys serve as data collection tools, providing researchers, businesses, and policymakers with access to valuable insights into human behavior, opinions, and preferences. They play a key role in market research, empowering businesses to understand consumer needs, preferences, and market dynamics. By eliciting feedback on products, services, and brand perceptions, surveys inform strategic decision-making, product development, and marketing strategies, driving competitive advantage and customer satisfaction. In the field of social research, surveys serve as instruments to measure public opinion, assess social attitudes, and understand demographic trends [1]. Whether in politics, education, or health care, surveys provide invaluable insights into the needs, perceptions, and behaviors of diverse populations, informing policy formulation, program development, and resource allocation. However, it is difficult to have a survey without missing data, which is caused by various reasons. Thus, before analyzing the gathered data, it is important to analyze the missing data, which can be a considerable part of the total [2].

Generally, in statistical analysis and processing, researchers often exclude cases with partial data. As an alternative to this method of data treatment, researchers can choose to replace some of the missing values of a variable with reliable values such as the average of the observed values for that variable, the median, or the average of the neighboring values of the missing data [3].

Some of the methods of replacing or completing missing data are already present in many popular statistical software programs, such as SPSS. In addition to deterministic techniques for filling in missing data, this software also includes stochastic techniques based on probabilistic data distribution models, such as the maximum likelihood technique, multiple imputations [4].

Missing values are endemic across the social sciences and family studies. About 50% of the participants in political or social survey data have missing values; many of the major data sets that are utilized in articles appearing in family journals have serious problems with missing values. A missing rate of 15% to 20% is common in educational and psychological studies [5].

The seriousness of the missing values depends in part on the amount of missing data, the missing data model, and the mechanism underlying the missing data. The model, quantity, and mechanism of missing values have significant effects on the outcome of a study [6].

The percentage of missing data is directly related to the quality of the statistical conclusions. The percentage of missing data significantly affects the quality of the analysis and statistical conclusions. Various estimates are made about the percentage of missing data that can be deleted without causing seriously biased results. One estimate is that if the missing data is less than 5% of all the data, then list deletion can be used; a missing rate of less than 10% is inconsequential, and a missing data rate of more than 10% will produce biased statistical results. However, the amount of missing data is not the sole factor that affects the statistical results [7].

The missing data mechanisms and the missing data patterns have a greater impact on research results than the proportion of missing data [8]. However, there is always the question of what to do with missing data in research studies and how to assess its impact on the study results.

Numerous articles and studies have been conducted in various research fields, such as sociology, political science, psychology, education, communications, oil industry production data, oilfield reservoir properties, tourism customer satisfaction, urban resilience criteria, academic performance, and teaching, through the use of massive data [9]-[10].

The problem of missing data is often present, so handling the "missingness" by implementing the most suitable imputation methods is necessary and advisable [11]-[12].

They have analyzed problems created by missing data and implemented different methods of data imputation, including deterministic (simple calculation) and probabilistic (value estimation) [13].

The problem of missing data has been addressed in several types of research in Albanian research studies: familiar, social, academic, economic, tourism services, health services, and industrial projects [14]-[15].

Although there is an extensive literature on missing data, social science researchers have been slow to adopt these methods or heed their recommendations [16]-[17].

The lack of taking missing data into consideration and finding and implementing the most appropriate techniques to complete them risks obtaining biased and distorted results and wrong conclusions from these surveys [18].

To begin with, a safer way is to review the stages of data collection and complete them by redoing the survey, but this is often not a convenient thing because the time of the survey is over, the people who have been asked cannot be more available, etc. However, the analysis of the stages of preparation and completion

of a questionnaire will highlight problems and shortcomings that will be taken into account in future cases [19].
The prevention of missing data is the most optimal approach to conducting a successful survey and reaching accurate and correct results and conclusions for the issue being studied [20]. During the data collection phase, researchers have the opportunity to decide what data to collect and how to monitor their collection [21].
The extent and distribution of variables in the data, along with the reasons for missing data, are critical factors in applying appropriate missing data techniques [22].
Missing data strategies, from complete case analysis to model-based methods, each rely on assumptions about the nature of the mechanism that causes missing data [23]. Before applying any imputation methods to replace missing data, the researcher must first diagnose and understand the missing data processes underlying the missing data [24].

Causes of Missing Data:
Data may be missing for various reasons, which may be objective or subjective. The reasons that data is missing can affect the appropriateness and value of the strategies used to address the problem. Missing data in surveys arises from various sources, each presenting unique challenges [25].
The specific causes of missing data can be due to specific characteristics (e.g., income level, political affiliation), which can significantly influence its impact on the research findings, introduce bias, potentially skewing the results and undermining the generalizability of the study [26].
Some of the most common causes include:

- **Non-response:** This occurs when individuals

refuse to participate in the survey or fail to answer specific questions. Reasons for nonresponse can range from lack of interest, privacy concerns, or difficulty understanding the questions.

- **Item nonresponse:** Individuals may participate

in the survey but skip specific questions due to sensitivity, lack of knowledge, or missing information.

- **Measurement errors:** These occur when

responses are recorded inaccurately or when questions are ambiguous, leading to missing or unusable data.

Consequences of Missing Data:
Missing data can pose significant threats to the validity and reliability of research findings. Here are some key consequences:

- **Bias:** Missing data can lead to biased estimates,

as individuals who do not respond or skip certain questions may differ systematically from those who do provide complete data. This bias can distort conclusions and hinder the drawing of accurate inferences.

- **Reduced power:** Incomplete data reduces the

sample size available for analysis, diminishing the statistical power of the study. This can make it difficult to detect statistically significant relationships and increase the risk of Type II errors (failing to reject a false null hypothesis).

- **Loss of information:** Missing data represents

valuable information left untapped, potentially hindering the understanding of complex phenomena and limiting the scope of research conclusions.

Classification of missing data:
Rubin introduced three types of missing data which are [27]:
Missing Completely at Random (MCAR):
Data is considered missing completely at random when there is no systematic underlying process as to why individuals are missing for a given measurement. It may be the case that some of the questionnaires were accidentally dropped for one or a few participants, or that another person was momentarily distracted by other things unrelated to the question or the answer or random technical errors. In this case, the missing data is unrelated to both the missing and observed values in the dataset;
$p(R/D_{mis}, D_{obs}) = p(R)$

Missing at Random (MAR):

Data is considered MAR if it is missing because of some potentially observable, non-random, systematic process. Data is considered MAR if the probability of missing data for some variable (Y) is predictable based on the value of another variable or set of variables (X). The missing data depends only on observed values in the dataset;

$p(R/D_{mis}, D_{obs}) = p(R/D_{obs})$

Missing Not at Random (MNAR):

If data is missing due to the value of the variable being considered, then it is not considered missing at random. If we are considering a variable Y, it would be MNAR if individuals chose not to respond because of the value of Y;

$p(R/D_{mis}, D_{obs}) \neq p(R/D_{obs})$

A classic example is the questionnaire about income. Income may often be MNAR because people who have a very high or very low income might choose not to report the value of their income because they do not feel comfortable with the numbers they have to declare. In this case, the missing data of the income variable is dependent upon the value of the variable, figure 1-3, (red is missing data in the *y*-variable, and blue is observed data).

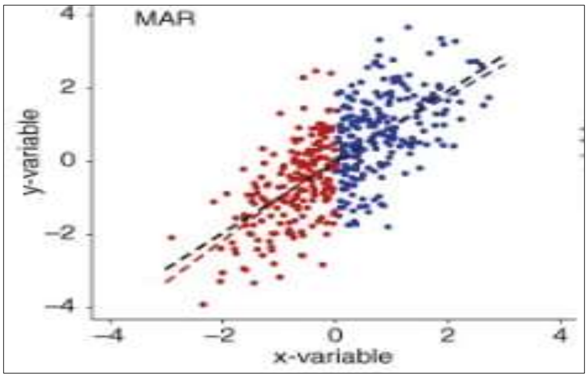

Fig. 1. Mechanism of missing data. MAR.

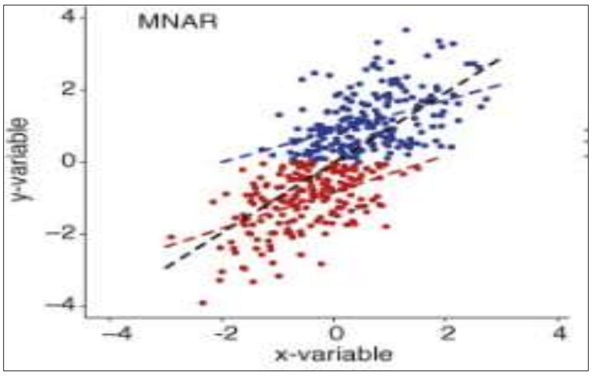

Fig. 2. Mechanism of missing data. MNAR

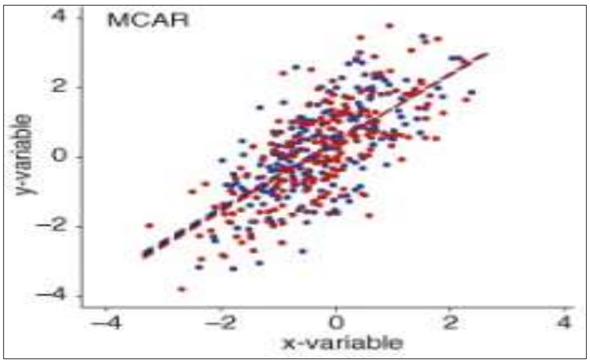

Fig. 3. Mechanism of missing data. MCAR

Handling the Missing Data:

The best method is not to have missing data [28]. Fortunately, researchers have developed various strategies to address missing data issues. These strategies can be broadly categorized into two approaches [29]:

- **Prevention:** Preventing missing data in the

first place is the ideal solution. This involves careful survey design, clear and concise questions, effective recruitment methods, and addressing potential barriers to participation.

- **Imputation:** When missing data occurs,

imputation techniques aim to estimate missing values based on available data. Common methods include mean imputation, regression imputation, and more sophisticated techniques like multiple imputation. It's crucial to choose the appropriate imputation method based on the type of missing data, underlying assumptions, and characteristics of the data.

Choosing the Right Approach:

The optimal approach for handling missing data depends on several factors, including the nature of the missing data, the research question, and the resources available. Careful consideration of these factors and expert guidance are essential to ensuring the chosen method does not introduce bias or further distort the data. The Little's chi-square statistic test serves to reject or not the Ho Hypothesis that the missing data are MCAR at a significance of 0.05 level. If the $p<.5$, then the null hypothesis is rejected, and the data are not MCAR. In that case, the data may be missing at random (MAR) or not missing at random (NMAR). If $p >.5$, then the null hypothesis is not rejected, so the data are missing completely at random (MCAR) [30].

Listwise deletion or case deleting:

Among the studies that showed evidence of missing data, 97% used the list deletion (LD) method or the pairwise deletion (PD) method to deal with missing data, which are well-known methods for biased and/or ineffective assessments in most statistical studies [31].

The most common method and the easiest way to deal with missing data that is used by most researchers is the list-wise method, which means deleting all the cases where there is at least one missing data point. Listwise deletion is an easy and simple method to implement because it is the default used in many statistical packages, including SPSS.

If the data is MCAR then the listwise deletion will not introduce any bias into parameter estimates, because, under MCAR, the subsample of cases with complete data is equivalent to a simple random sample from the original target sample. The obvious downside of listwise deletion is that it often discards a great deal of potentially usable data which leads to larger standard errors, wider confidence intervals, and a loss of power in testing hypotheses. Listwise deletion typically results in the loss of 20%–50% of the data [32].

Pairwise deletion:

Pairwise deletion is the case when we delete only the missing data in the present variables. In the end, some variables may have a different number of cases compared to other variables.

Pairwise deletion is, for linear models, a very popular alternative replacement for listwise deletion.

Like listwise deletion, however, if the data are MAR but not MCAR, the pairwise deletion may produce biased estimates. Pairwise deletion may be more efficient than listwise deletion because more data is utilized in producing the estimates [33].

Imputation methods:

Such methods produce some estimation for each missing value by using the proper software [34].

Mean imputation is the simplest and most popular approach for missing values, but this is well known to produce biased estimates [35].

The mean is calculated for all the non-missing data variables, and all the missing data will be substituted with the mean values.

It has the advantage of keeping the same mean and the same sample size, but many disadvantages, such as the false impression of sample size and the decreasing of the variance.

If the purpose of the research is mean estimations or if the data are missing completely at random, the series mean imputation will not bias the parameter estimate; it will still bias the standard error.

The mean of nearby points replaces missing values with the mean of valid surrounding values (2 or more). The span of nearby points is the number of valid values above and below the missing value used to compute the mean.

Median of nearby points replaces missing values with the median of valid surrounding values (2 or more). The span of nearby points is the number of valid values above and below the missing value, which are used to compute the median for the purpose of substituting the missing values.

Linear interpolation replaces missing values using linear interpolation. The last valid value before the missing value and the first valid value after the missing value are used for the interpolation.

The "linear trend at point" method essentially performs a regression where the variable with missing values is the dependent variable and the case sequence number is the predictor.

One problem is that they can produce biased estimates of some parameters. In particular, variances for missing data variables tend to be underestimated, along with any parameter affected by variances (e.g., regression coefficients).

The EM (Expectation-Maximization) method assumes a distribution for the partially missing data and bases inferences on the likelihood under that distribution [36].

Each iteration consists of an E step and an M step. The E step finds the conditional expectation of the "missing" data, given the observed values and current estimates of the parameters.

These expectations are then substituted for the "missing" data. In the M step, maximum likelihood estimates of the parameters are computed as though the missing data had been filled in.

II. MATERIALS AND METHODS

The Likert scale is a popular psychometric scale implemented for measuring attitudes, opinions, and perceptions on many topics [37].

On a 5-point Likert scale, respondents are typically asked to indicate their level of agreement or disagreement with a statement, with the following options:

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree (or neutral)
4. Agree 5. Strongly agree

This scale allows respondents to express their opinions along a continuum from strong disagreement to strong agreement, providing researchers with ordinal data that can be analyzed quantitatively.

It's a versatile tool used in various fields such as psychology, sociology, market research, and more. The survey contains 13 questions about tourism services in Albania during the tourist season, June–October 2023. All 146 forms were filled with answers, out of 180, Table 1.

Table 1. Evaluation form with 13 questions

Nr.	Questions to evaluate	1	2	3	4	5
1	Tourist attractions are well-maintained.					
2	The local people are friendly and welcoming.					
3	The transportation are convenient and reliable.					
4	The accommodations meet my expectations.					
5	The information provided is accurate and useful:					
6	The natural beauty of this destination is appealing.					
7	The local cuisine and dining options are enjoyable.					
8	The cultural and historical heritage are interesting.					
9	The overall safety and security are satisfactory.					
10	The cost of tourism activities and services is reasonable.					
11	The variety and quality of available activities and attractions are satisfactory.					
12	The overall value for money of the tourism services is satisfactory:					
13	My overall satisfaction with the tourism services during is satisfactory:					

III. RESULTS

The SPSS software is implemented to calculate univariate Statistics, table 2. The Little' test of MCAR is conducted with EM means. The result is p= Sig. = .119> .05, meaning that the missing data are MCAR, (Missing Completely at Random), table 3.
The means of survey's variables are included in the interval of values (3.44; 4.02) which is interpreted as "agree.", table 4. This level of satisfaction for all the variables means that generally, tourists are content with the services.

Table 2. Survey's data Statistics, with SPSS

Univariate Statistics							
				Missing		No. of Extremes [a]	
	N	Mean	Std. Deviation	Count	Percent	Low	High
Accommodation	137	3.67	.823	9	6.2	1	0
Information	139	3.69	.849	7	4.8	0	0
Destination	139	3.94	.778	7	4.8	0	0
Cuisine	138	3.76	.882	8	5.5	0	0
Heritage	140	3.83	.725	6	4.1	0	0
Safety	139	3.90	.741	7	4.8	0	0
Cost	141	3.55	.814	5	3.4	3	0
Quality	136	3.75	.696	10	6.8	0	0
Value	140	3.65	.879	6	4.1	0	0
Overall	141	3.80	.852	5	3.4	2	0
Attraction	140	3.65	.821	6	4.1	0	0
Atmosphere	139	4.02	.746	7	4.8	3	0
Transportation	137	3.44	.812	9	6.2	1	0
a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).							

Table 3. Little's test of MCAR

Mean value	Interpretation
1.00- 1.8	Strongly disagree (SD)
1.81- 2.6	Disagree (D)
2.61- 3.4	Neutral (N)
3.41- 4.2	Agree (A)
4.21- 5.0	Strongly Agree (SA)

Table 4. Interval of mean values and interpretation

EM Means[a]												
1	2	3	4	5	6	7	8	9	10	11	12	13
3.6403	3.6864	3.9486	3.7702	3.8413	3.8997	3.5535	3.7414	3.6566	3.8146	3.6563	4.0152	3.4380
a. Little's MCAR test: Chi-Square = 256.561, DF = 231, Sig. = .119												

By default, the SPSS software listwise deletion reduces the partial cases (with missing data) to 111 from a total of 146, or 24% of all the cases, which is a considerable amount, table 5.
Such a large percentage may be a good reason to search suitable methods of data imputation. The overall missing data (%) and the pattern of missing data are shown in figures 4 and 5.
The Cronbach's Alpha, which is a measure of internal consistency, is calculated as CA =.801 which is interpreted as "very good", table 6.
The data imputation is implemented with the EM (Expectation-Maximization) method), the results of EM calculation are different from the listwise results, table 7.

Table 5. listwise deletion (%)

Case Processing Summary			
		N	%
Cases	Valid	111	76.0
	Excluded [a]	35	24.0
	Total	146	100.0
a. Listwise deletion based on all variables in the procedure.			

Table 6. Cronbach's test of consistency

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.799	.801	13

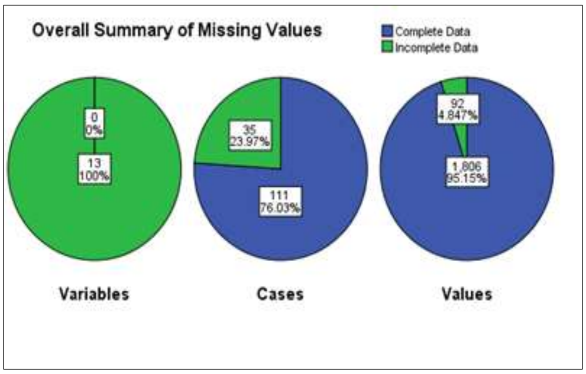

Fig. 4. Overall summary of missing data (%)

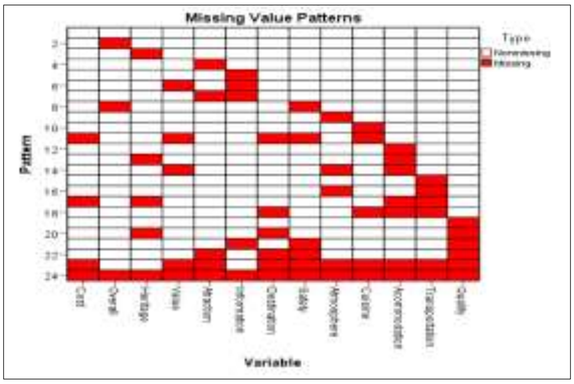

Fig. 5. Pattern of missing data

Table 7. Listwise, pairwise, EM and Regression estimated means

Summary of Estimated Means													
	1	2	3	4	5	6	7	8	9	10	11	12	13
Listwise	3.72	3.74	3.97	3.80	3.85	3.90	3.59	3.77	3.75	3.83	3.78	3.99	3.48
All Values	3.67	3.69	3.94	3.77	3.84	3.91	3.55	3.75	3.66	3.81	3.65	4.02	3.45
EM	3.64	3.69	3.95	3.77	3.84	3.90	3.55	3.74	3.66	3.81	3.66	4.02	3.44
Regression	3.68	3.67	3.95	3.77	3.86	3.91	3.56	3.75	3.66	3.82	3.67	4.02	3.42

IV. CONCLUSİON

The questionnaire was created to evaluate the tourism services of several sites in Durres and Berat, two important touristic destinations with different characters.

Durres and its regions are the main tourist destinations on the Albanian Adriatic coast. They are characterized by massive tourism for summer vacations by locals and foreigners.

On the other hand, Berat is a cultural elite tourist destination throughout the year due to the archaeological-cultural heritage inherited from more than 2000 years of history and tradition. The city has a great attraction for tourists interested in history: the ancient castle, the characteristic old neighborhoods, the Mediterranean cuisine, and the beautiful nature that surrounds it, with the high mountains in front, the green hills that surround it, and the river that runs through.

During the survey process, most of the forms were filled in properly; still, many forms were not filled in at all, several were found not suitable to consider, other forms were not filled in properly, and there were many forms with missing answers. Forms were distributed to foreign tourists for their voluntary cooperation.

We estimate that there are no questions that require any information that could create uncertainty or avoidance of answers. We believe that the missing values were completely accidental and have nothing to do with answers, so they are generally MCAR.

The process produced much missing data, and for that reason, an analysis of the missing problem was implemented and data imputation methods were evaluated.

Generally, missing data in surveys can arise from a multitude of factors and pose significant challenges for researchers. Its implications extend beyond mere data incompleteness, impacting the validity, reliability, and generalizability of research findings. Missing data in surveys remains a complex challenge, but with careful planning, appropriate strategies, and ongoing advancements in statistical methods, researchers can navigate this labyrinth and obtain reliable and insightful findings from their surveys.

The simplest way to fill in the missing data is to eliminate all the incomplete cases. This method is included in most statistical programs, such as SPSS, R, etc. However, in cases where the missing data are MAR or NMAR, or their percentage is more than 10%, this method will cause biased results.

Several techniques and methods are implemented for missing data imputation, such as mean substitutions, median substitutions, linear regression, the mean of unknown points, Considering the EM probability imputation method as the most reliable for estimating missing data, the result shows that other methods produce different values, which may have a significant impact on the estimation of oil and gas reserves for the three valuation scenarios.

ACKNOWLEDGMENT

The survey of tourism services in Albania is part of the ongoing project "MCDA to evaluate local and national projects with the help of multicriteria decision analyses", sponsored by the University "Aleksander Moisiu" Durrës, Albania.

REFERENCES

[1] Schneider, C. R. (2022). Missing data in surveys: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, *18*(2), 2308-2316.

[2] Cook, R. M. (2021). Addressing missing data in quantitative counseling research. Counseling Outcome Research and Evaluation, 12(1), 43-53.

[3] Firat, M., Dikbas, F., Koç, A. C., & Gungor, M. (2010). Missing data analysis and homogeneity test for Turkish precipitation series. *Sadhana*, *35*(6), 707-720.

[4] Halidini, D., Xhafaj, E., & Gjikaj, N. Treating the missing values for the total waste recycling in Albania. Interdisiplinary Journal of Research and Development, Vol. 4, no. 1. 2017

[5] Acock, A. C. (2005). Working with missing values. *Journal of Marriage and family*, *67*(4), 1012-1028.

[6] Kosova, R., Xhafaj, E., Karriqi, A., Boci, B., & Guxholli, D. (2022). Missing Data In The Oil Industry And Methods Of Imputations Using Spss: The Impact On Reserve Estimation. *Journal of Multidisciplinary Engineering Science and Technology*, *9*(2), 15146-15155.

[7] Lin, W. C., & Tsai, C. F. (2020). Missing value imputation: a review and analysis of the literature (2006–2017). *Artificial Intelligence Review*, *53*, 1487-1509.

[8] Anaby-Tavor, A., Carmeli, B., Goldbraich, E., Kantor, A., Kour, G., Shlomov, S., ... & Zwerdling, N. (2020, April). Do not have enough data? Deep learning to the rescue!. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 05, pp. 7383-7390).

[9] Scarelli, A., Kosova, R., & Sinaj, V. (2018). Multicriteria decision aid for measuring resilience in a three-dimensional space. DOI 10.33107/ubt-ic.2018.97

[10] Kosova, R., Shehu, V., Naco, A., Xhafaj, E., Stana, A., & Ymeri, A. (2015). Monte Carlo simulation for estimating geologic oil reserves. A case study from Kucova Oilfield in Albania. Muzeul Olteniei Craiova. Oltenia. Studii şi comunicări. Ştiinţele Naturii, 31(2), 20-25. ISSN 1454-6914.

[11] Kosova, R., Naco, A., & Prifti, I. (2016). Deterministic and Stochastic methods of oil field reserves estimation: A case study from KA. Oil field. *Interdisplinary Journal of Research and Development*, *4*(2), 226-231.

[12] Kosova, R., Subashi, D. X., Halidini, D. Q., & Kullolli, T. (2023). MCDA for evaluating the competitiveness of tourist destinations in Albania. *International Journal of Advanced Natural Sciences and Engineering Researches (IJANSER)*, *7*(11), 95–102. https://doi.org/10.59287/as-ijanser.569

[13] Kosova, R., Halidini, D. Q., Xhafaj, E., Gjikaj, N., & Kosova, A. M. (2023). Assessing tourism service quality in Albania: A 5-scale Likert survey analysis and interpretation with SPSS. http://dx.doi.org/10.59287/as-proceedings.367

[14] Bici, R. (2023). Simple methods to handle missing data. *International Journal of Computational Economics and Econometrics*, *13*(2), 216-242.

[15] Kosova, R., Qendraj, D. H., & Xhafaj, E. (2022). Meta-Analysis ELECTRE III and AHP in Evaluating and Ranking the Urban Resilience. *Journal of Environmental Management & Tourism*, *13*(3), 756-768. DOI http://dx.doi.org/10.14505/jemt.v13.3(59).15

[16] Gjana, A., & Kosova, R. Traditional Class, and Online Class Teaching. Comparing the Students Performance Using ANCOVA. *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, 14806-14811.

[17] Kosova, R., & Sinaj, V. (2021). Mathematical modeling of tourism development. An application to albanian tourism. *Journal of Environmental Management & Tourism*, *12*(6 (54)), 1707-1715. http://dx.doi.org/10.14505//jemt.v12.6(54).26

[18] Kosova, R., & Prifti, I. (2021). Missing data in the oil industry. Method of imputations and the impact on reserve estimation. In *Conference on Applied and Industrial Mathematics* (pp. 28-29).

[19] Kosova, R., & Sinaj, V. (2020). SERVICE QUALITY AND HOTEL CUSTOMER SATISFACTION: A CASE STUDY FROM DURRES, ALBANIA. *Annals Series, of' Constantin Brancusi'University of Targu-Jiu. Economy* (6).

[20] Kosova, A. G. R. The Performance Of University Students And High School Factors. Statistical Analyses And ANCOVA.

[21] Kosova, R., Sinaj, V., Scarelli, A., Kosova, A. M., & Stana, A. (2017). Academic staff performance evaluation, using decision analyses method and application. http://dx.doi.org/10.33107/ubt-ic.2017.124

[22] Bruni, R., Daraio, C., & Aureli, D. (2021). Imputation techniques for the reconstruction of missing interconnected data from higher educational institutions. *Knowledge-Based Systems*, *212*, 106512.

[23] Zdebska, W. (2021). The occurrence of missing data in surveys. *Acta Scientiarum Polonorum. Oeconomia*, *20*(2), 95-103.

[24] KAYA, M., & Saleem, A. M. (2023). Predictive Analysis in E-commerce: Utilizing Data Mining Techniques to Forecast Customer Purchasing Behavior. *International Journal of Advanced Natural Sciences and Engineering Researches (IJANSER)*, *7*(11), 194–200. https://doi.org/10.59287/as-ijanser.583

[25] Hajrulla, S., Hajrulla, D., Lino, V., & Stojani, T. (2023). Statistical Data on Risk Management in University. Overview of Numerical Results. *International Journal of Advanced Natural Sciences and Engineering Researches (IJANSER)*, *7*(11), 13–19. https://doi.org/10.59287/as-ijanser.558

[26] Demir, T., Hajrulla, S., Bezati, L., & Hajrulla, D. (2023). Application of Numerical Methods in Design of Industrial Structures. *International Journal of Advanced Natural Sciences and Engineering Researches (IJANSER)*, *7*(11), 295–300. https://doi.org/10.59287/as-ijanser.632

[27] Akif, A. V. C. U. (2021). Effects of Different Multiple Imptutation Techniques on the Model Fit of Confirmatory Factor Analysis. *Trakya Eğitim Dergisi*, *11*(3), 1227-1238.

[28] Fisher, R. A. (1925, July). Theory of statistical estimation. In *Mathematical proceedings of the Cambridge philosophical society* (Vol. 22, No. 5, pp. 700-725). Cambridge University Press.

[29] Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

[30] Little, R. J. (2021). Missing data assumptions. *Annual Review of Statistics and Its Application*, *8*, 89-107.

[31] Grund, S., Lüdtke, O., & Robitzsch, A. (2021). On the treatment of missing data in background questionnaires in educational large-scale assessments: An evaluation of different procedures. *Journal of Educational and Behavioral Statistics*, *46*(4), 430-465.

[32] Nakagawa, S., & Freckleton, R. P. (2008). Missing in action: the dangers of ignoring missing data. Trends in ecology & evolution, 23(11), 592-596.

[33] Howell, D. C. (2007). The treatment of missing data.

[34] Tasho, E. M., & Kafazi, D. Using multiple imputation methods to analyze missing values for real and simulated data. In *Book of proceedings* (p. 11).

[35] Barnum, M. (2022). Dealing with missing and incomplete data. In *Handbook of Research Methods in International Relations* (pp. 425-445). Edward Elgar Publishing.

[36] Alruhaymi, A. Z., & Kim, C. J. (2021). Study on the Missing Data Mechanisms and Imputation Methods. *Open Journal of Statistics*, *11*(4), 477-492.

[37] Jebb, A. T., Ng, V., & Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Frontiers in psychology*, *12*, 637547.