Analysis of Geographical effect of various regions on Liver disease

Bendi Venkata Ramana
Department of IT, Aditya Institute of Technology and Management, Tekkali, A.P, India
Email: ramana.bendi@gmail.com

ABSTRACT

Statistical Analysis plays a significant role in population comparison to investigate the geographical effect on liver diseases. In this study the common attributes ALP, DB, SGOT, SGPT and TB were considered from the three datasets for the population comparison. Three data sets were assessed using analysis of variance and multivariate analysis of variance and significance level observed for the statistical analysis is <= 0.05 for the corresponding confidence level is 95%. The Significant values in the ANOVA and MANOVA analysis indicates there is more significant difference among three liver datasets that means there is a geographical effect on liver diseases.

Keywords

Statistical Analysis, liver datasets, population comparison

Introduction

Statistical Analysis plays a significant role in population comparison to investigate the geographical effect on liver diseases. In this study the common attributes were considered from the three datasets for the population comparison. Three data sets were assessed using analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA). ANOVA analysis applied on each single dependent variable that is ALP, DB, SGOT, SGPT and TB to test the significant difference among three datasets. MANOVA statistical analysis [1-5] applied on all combinations of common attributes to envisage the significance among three datasets.

Related Work

Wong et al. evaluated hepatic fibrosis among Nonalcoholic fatty liver disease patients, concentrating on dietary patterns [6]. Xia et al. investigated rate of Nonalcoholic fatty liver disease regional wise among liver fat content and glucose metabolism [7]. Cannon et al. justified exceptions other than hepatocellular carcinoma for final stage liver disease [8]. Jothimani et al. reported that the mortality rate is more for liver patients with COVID-19 [9]. Author Liao et al. explored the correlation among medical-level factors, county-level factors and rate of antibiotic use [10]. Khorraminezhad et al. explored the approaches for the analysis of multi-OMICs data in nutrition studies [11]. Bendi, V.R. et al. investigated performance of selected classification algorithms on medical datasets taken from the UCI repository [12]. Percival et al. analyzed statistically using multicomponent nuclear magnetic resonance [13]. Yousaf et al. forecasted COVID-19 confirmed cases and death cases using the Auto-Regressive Integrated Moving Average Model [14]. Ustaoglu et al. accomplished statistical analysis using pearson correlation index, principal component and clustering analysis to determine water quality of the stream [15]. Petersen et al. performed statistical analysis on many patients increasing hypoxic respiratory failure due to COVID-19 [16]. Gyimah et al. investigated statistically using principal component and cluster analyses for the assessment of Densu River water quality [17]. Kattan et al. explored papers that are submitted to CHEST will be reviewed using the statistical analysis and give response to peer review [18]. Ren et al. accomplished statistical analysis related to tunnel fire accidents based on causes, characteristics, and consequences [19]. Author Livadiotis implemented a statistical analysis on progress rate of COVID-19 cases with the effect of temperature [20]. Verma et al. analyzed statistically using KNN and K-means on CIDDS-001 dataset for assessing Network Intrusion Detection Systems [21]. Zajkowska et al. performed multivariate statistical analysis for diagnosing breast cancer at an initial and final stage [22]. Mustaqeem et al. developed a model that predicts cardiac disease patients by assessing clinical features of patients [23]. Zhao et al. examined transmission chain of secondary cases COVID-19 [24]. Lara-Cabrera et al. considered the efficacy of metrics to assess the risk of radicalization [25]. Barbulescu et al. considered the perceptions dependent on the intensity of dust storm every month, season, year and location [26]. Sidorov et al. deliberated the practical approach for processing the research data by using research techniques and statistics [27]. Nordhausen et al. anticipated Independent component analysis for the refinement of principal component analysis by using covariance matrix [28].

Data Sets

Datasets were taken from Visakhapatnam, Vijayawada and Tirupathi based on the major geographical regions of Andhra Pradesh that are North Coastal Andhra Pradesh, Central Andhra Pradesh and Rayalaseema respectively. These datasets [21] are analyzed by using statistical techniques to know the impact of geographical variables such as food habits, behaviors, environment etc on Liver Function Tests (LFT) [22-23]. The three datasets are named Visakhapatnam dataset, Vijayawada dataset and Tirupathi dataset based on geographical region. Visakhapatnam dataset contains 12 attributes and has 499 samples. Vijayawada dataset contains 12 attributes and has 600 samples. Tirupathi dataset contains 7 attributes and has 243 samples. The description of datasets is given table1.

Table 1. Description of datasets from different regions of Andhra Pradesh

Datasets	# Attributes	# Samples	# Classes
Visakhapatnam Dataset	7	243	2
Vijayawada Dataset	12	600	2
Tirupathi Dataset	12	499	2

Statistical Analysis

Statistical Analysis involves the investigation of techniques for gathering, summing up, and deciphering information. Measurements formalize the way toward deciding. The uses of measurements in sciences, financial aspects, software engineering, money, brain research, humanism, criminology, and numerous different fields. It will inspect various approaches to research the connections between different attributes of information. In this investigation standard statistical techniques [25] ANOVA and MANOVA are applied to assess the importance between two populaces for better grouping. ANOVA is utilized to test the considerable difference between one dependent variable and one independent variable. MANOVA is utilized to test the considerable difference between more dependent variable and individual independent variable. The truth liver datasets were taken from Visakhapatnam, Vijayawada and Tirupathi based on the main geographical regions of Andhra Pradesh that are North Coastal Andhra Pradesh, Central Andhra Pradesh and Rayalaseema region respectively. The attributes in these datasets are represented in the table 2. The common attributes from the three data sets AGE, GENDER, TB, DB, SGOT, SGPT and ALP are considered for the purpose of population comparison to analyze the geographical effect. In this Group 1 indicates Visakhapatnam dataset, Group 2 indicates Vijayawada dataset and Group 3 indicates Tirupathi dataset.

Table 2. List of attributes and types of regional datasets

Visakhapatnam Region		*Vijayawada Region*		*Tirupathi Region*	
Attribute	***Type***	***Attribute***	***Type***	***Attribute***	***Type***
AGE	Real number	**AGE**	Real number	**AGE**	Real number
GENDER	Categorical	**GENDER**	Categorical	**GENDER**	Categorical
TB	Real number	**TB**	Real number	**TB**	Real number
DB	Real number	**DB**	Real number	**DB**	Real number
SGOT	Integer	**AST (SGOT)**	Integer	**SGOT**	Integer
SGPT	Integer	**ALT (SGPT)**	Integer	**SGPT**	Integer
ALP	Integer	**ALP**	Integer	**ALP**	Integer
		IB	Real number	**IB**	Real number
		SP (TP)	Real number	**TP**	Real number
		SA (Albumin)	Real number	**Albumin**	Real number
		Globulins	Real number	**Globulins**	Real number
		A/G RATIO	Real number	**A/G RATIO**	Real number

Analysis of variance

ANOVA is used to analyze statistically [14-16] between one dependent variable and two or more categories. In this between groups directs variability based on place of data, within groups directs variability based on random error and total directs total variability. F-statistic is a ratio of variation between group and within group. ANOVA and MANOVA statistical techniques are considered for the population comparison between datasets to know the geographical effect on liver datasets. Statistical analysis metrics depicted in Fig. 1.

$$SS_{between} = n\sum(\bar{x} - \bar{x}_{grand})^2$$

$$SS_{within} = \sum(x - \bar{x}_{group})^2$$

$$F - statistic = \frac{\text{Between Group Variation}}{\text{Within Group Variation}}$$

$$X_{grand\ mean} = \frac{\sum \bar{x}}{N}$$

$$X_{group\ mean} = \frac{\sum fm}{\sum f}$$

$SS_{between}$ = Sum of the squares between groups

SS_{within} = Sum of the squares within Groups

N = Total number of sets

$\sum \bar{x}$ = Sum of mean of all sets

$df_{total} = N - 1$

$df_{between} = K - 1$

$df_{within} = N - K$

N = Number of all scores

K = Number of groups

Degrees of freedom = number of values estimated by the statistical analysis

f = frequency

m = midpoint

X = Sample value

$\bar{x}$ = Mean of the sample

Fig. 1 Statistical Analysis Metrics

Multivariate Analysis of variance

MANOVA is utilized to assess the hypothesis that at least one independent factors (IVs), affect a group of at least two dependent factors (DVs). The objective of our investigation is to search for an impact of at least one IVs on a few DVs simultaneously. Four distinctive multivariate tests [17] were considered to distinguish the huge impact of the IVs on the entirety of the DVs, as a gathering. Descriptive statistics are categorized into two measures. One is central tendency that comprise mean, median, and mode. Another one is dispersion of data that comprises standard deviation, variance, the minimum and maximum.

Results and Discussion

The common features ALP, DB, SGOT, SGPT and TB were considered as DVs and group was considered as factor variable. The consequences of ANOVA were addressed in three lines. The outcome of ANOVA analysis on ALP, DB, SGOT, SGPT and TB was addressed in table 3, table 4, table 5, table 6 and table 7 correspondingly which demonstrates whether there is a measurably difference among means of groups.

Table. 3: ANOVA on ALP between three datasets

ANOVA - ALP	***Sum of Squares***	***df***	***Mean Square***	***F***	***Sig.***
Between Groups	382.012	226	1.690	6.057	.000
Within Groups	311.154	1115	.279		
Total	693.165	1341			

Table. 4: ANOVA on DB between three datasets					
ANOVA - DB	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Between Groups	279.482	46	6.076	19.019	.000
Within Groups	413.684	1295	.319		
Total	693.165	1341			

Table. 5: ANOVA on SGOT between three datasets					
ANOVA - SGOT	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Between Groups	257.506	127	2.028	5.650	.000
Within Groups	435.659	1214	.359		
Total	693.165	1341			

Table. 6: ANOVA on SGPT between three datasets					
ANOVA - SGPT	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Between Groups	179.806	129	1.394	3.288	.000
Within Groups	513.323	1211	.424		
Total	693.129	1340			

Table. 7: ANOVA on TB between three datasets					
ANOVA - SGPT	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Between Groups	141.279	61	2.316	5.372	.000
Within Groups	551.886	1280	.431		
Total	693.165	1341			

Significance value indicates the possibility of getting a change in mean among the groups as high as seen by chance. More considerable change among the groups due to the lower p-value. In this investigation the sig. value is less than 0.05 means there will be more considerable variance among groups. This specifies the groups have considerable variances on ALP, DB, SGOT, SGPT and TB. ANOVA explores statistically considerable change totally among the groups and there is need of MANOVA analysis for Multiple Comparisons, to explore which groups differed from each other. Descriptive Statistics of Visakhapatnam dataset, descriptive Statistics of Vijayawada dataset and descriptive Statistics of Tirupathi dataset are represented in table 8, table 9 and table 10 correspondingly.

Table 8. Descriptive Statistics of Visakhapatnam Dataset				
Attribute	*Minimum*	*Maximum*	*Mean*	*Standard Deviation*
AGE	5	68	29.207	12.226
TB	0.3	18.2	1.474	2.172
DB	0.1	13	0.604	1.514
SGOT	14	175	40.626	26.324
SGPT	10	267	36.247	32.343
ALP	28	605	95.753	87.995

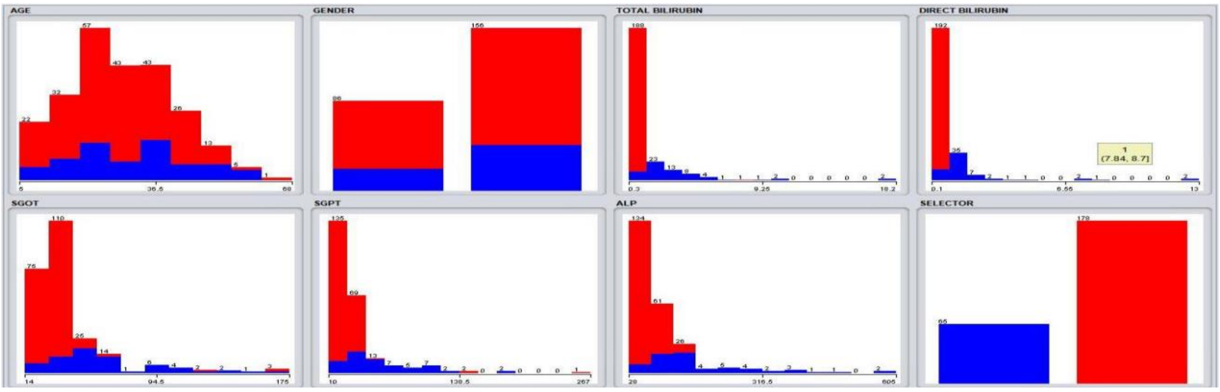

Fig. 1. Descriptive Statistics of Visakhapatnam Dataset

Table 9. Descriptive Statistics of Vijayawada Dataset				
Attribute	***Minimum***	***Maximum***	***Mean***	***Standard Deviation***
AGE	0	90	45.748	17.782
TB	0.1	31.6	0.925	1.886
DB	0.1	22.8	0.447	1.433
SGOT	2	4980	58.523	244.4
SGPT	2	3642	57.552	220.689
ALP	0.7	769	114.74	90.394
IB	0.1	15	0.511	0.979
TP	2.3	502	7.842	20.523
Albumin	0.8	4.9	3.147	.0754
Globulins	12	24	3.735	1.37
A/G RATIO	0.1	2.5	0.903	0.321

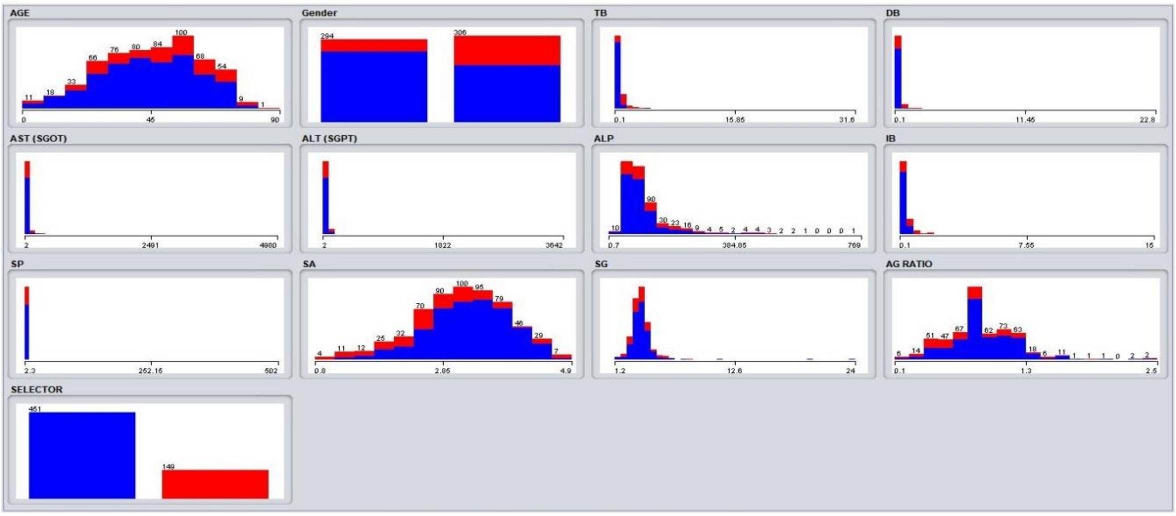

Fig. 2. Descriptive Statistics of Vijayawada Dataset

Table 10. Descriptive Statistics of Tirupathi Dataset				
Attribute	***Minimum***	***Maximum***	***Mean***	***Standard Deviation***
AGE	19	85	47.259	13.905
TB	0.2	1	0.726	0.138
DB	0.1	0.5	0.257	0.054
SGOT	10	45	23.733	7.97
SGPT	11	45	20.657	6.27
ALP	18	188	90.717	19.38

IB	0.1	0.7	0.468	0.128
TP	5.8	72	7.331	2.927
Albumin	2.9	4.8	4.17	0.293
Globulins	2.1	25	3.09	1.053
A/G RATIO	1	2.2	1.365	0.239

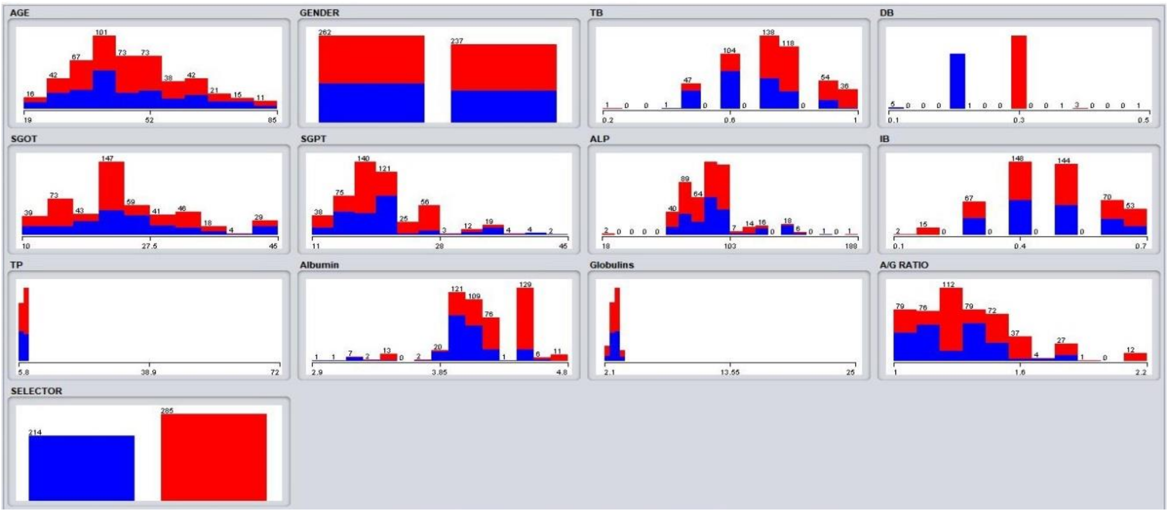

Fig. 3. Descriptive Statistics of Tirupathi Dataset

Multivariate analysis offers more substitute statistical tests and whereas one statistical test available in ANOVA. It has Wilks' Lambda, Lawley's trace, Pillai's trace and Roy's largest root statistical tests. They have generated same results with the hypothesis degrees of freedom is 1 and whereas for the degrees of freedom is greater than 1, Wilks' Lambda, Lawley's trace, and Roy's largest root are more powerful than Pillai's trace. MANOVA studies the degree of variance within the IVs and concludes the degree of variance between the IVs. IVs had a significant effect on the DV in within subjects variance and smaller than the between subjects variance. In this study the common attributes from the three datasets TB, DB, SGOT, SGPT and ALP are considered for the multivariate analysis. Multivariate assessments for the blend of attributes at the respective significant values at diverse levels are denoted in Table 11.

Table 11. Multivariate Tests Significance (at $p < 0.05$ level)

Level	*variables*	*Pillai's Trace Value*	*F*	*Hypothesis df*	*Error df*	*Sig.*	*Partial Eta Squared*
2-way Interactions	**DB*ALP**	.629	1133.53	2.000	1338.000	.000	.629
	DB*SGOT	.135	104.031	2.000	1338.000	.000	.135
	DB*SGPT	.137	106.212	2.000	1337.000	.000	.137
	SGOT*ALP	.628	1127.60	2.000	1338.000	.000	.628
	SGPT*ALP	.628	1126.30	2.000	1337.000	.000	.628
	SGOT*SGPT	.065	46.306	2.000	1337.000	.000	.065
	TB*ALP	.630	1138.87	2.000	1338.000	.000	.630
	TB*DB	.399	443.812	2.000	1338.000	.000	.399
	TB*SGOT	.286	268.204	2.000	1338.000	.000	.286
	TB*SGPT	.288	270.857	2.000	1337.000	.000	.288
3-way Interacti	**DB*SGOT*ALP**	.631	761.415	3.000	1337.000	.000	.631
	DB*SGOT*SGPT	.142	73.857	3.000	1336.000	.000	.142

ons	**SGOT*SGPT*ALP**	.628	751.625	3.000	1336.000	.000	.628
	TB*DB*SGOT	.407	306.048	3.000	1337.000	.000	.407
	TB*DB*SGPT	.410	308.974	3.000	1336.000	.000	.410
4-way Interactions	**DB*SGOT*SGPT* ALP**	.631	571.128	4.000	1335.000	.000	.631
	TB*DB*SGOT*ALP	.682	717.477	4.000	1336.000	.000	.682
	TB*SGOT*DB* SGPT	.392	215.767	4.000	1337.000	.000	.392
5-way Interactions	**TB*DB*SGOT* SGPT*ALP**	.683	574.015	5.000	1334.000	.000	.683

The combination of common attributes at different levels are 2-way relations, 3-way relations, 4-way relations and 5-way relations. The combination of attributes at 2-way relations are DB*ALP, DB*SGOT, DB*SGPT, SGOT*ALP, SGPT*ALP, SGOT*SGPT, TB*ALP, TB*DB, TB*SGOT and TB*SGPT. The combination of attributes at 3-way relations are DB*SGOT*ALP, DB*SGOT*SGPT, SGOT*SGPT*ALP, TB*DB*SGOT and TB*DB*SGPT. The combination of attributes at 4-way relations are DB*SGOT*SGPT*ALP, TB*DB*SGOT*ALP and TB*SGOT*DB*SGPT. The combination of attributes at 5-way Interaction is TB*DB*SGOT*SGPT*ALP. In this experimentation for the 95% of confidence level the equivalent significance level is 0.05. The Significant values are 0.000 for all the combinations of common attributes in multivariate analysis depicted in table 15 which is less than significant level i.e. 0.05. This indicates more significant difference exists between groups null hypothesis. For all the combinations that are DB*ALP, DB*SGOT, DB*SGPT, SGOT*ALP, SGPT*ALP, SGOT*SGPT, TB*ALP, TB*DB, TB*SGOT, TB*SGPT, DB*SGOT*ALP, DB*SGOT*SGPT, SGOT*SGPT*ALP, TB*DB*SGOT, TB*DB*SGPT, DB*SGOT*SGPT*ALP, TB*DB*SGOT*ALP, TB*SGOT*DB*SGPT and TB*DB*SGOT*SGPT*ALP they vary very much in population comparison.

Conclusions

The familiar attributes are ALP, DB, SGOT, SGPT and TB are considered from the three data sets for ANOVA and MANOVA. The significance level considered for the analysis is 0.05 with respect to confidence level is 95%. In this investigation ANOVA analysis shows more significant difference between groups due to rejection of null hypothesis and it these groups designates more difference on ALP, DB, SGOT, SGPT and TB. Simultaneously in MANOVA analysis there exists more significant difference due to the null hypothesis based on significant value 0.000. The results indicate that there is more significant difference among three liver datasets that means there is a geographical effect on liver diseases. This may be due to food habits, alcoholic consumption, air pollution, life style etc. Then there is a need of localized modifications for the identification of liver diseases.

Future Scope

ANOVA and MANOVA analysis is also suggested for the various confidence levels like 99 % and 90 %. This statistical analysis may be applied for various regions of India i.e different states of India to investigate the geographical effect and to suggest the localized settings for the diagnosis of liver diseases.

Acknowledgments

I am thankful to UGC, New Delhi for granting financial support under UGC Minor Research Project scheme with reference no. MRP-6934/16 (SERO/UGC). I am also thankful to Prof. V. V. Nageswara Rao, Director of Aditya Institute of Technology and Management (A) for providing necessary facilities.

References

[1] Nibedan Panda Sir: Panda, N. and Majhi, S.K., 2020. Improved spotted hyena optimizer with space transformational search for training pi-sigma higher order neural network. Computational Intelligence, 36(1), pp.320-350.

[2] Nibedan Panda Sir: Panda, N., Majhi, S.K., Singh, S. and Khanna, A., 2020. Oppositional spotted hyena optimizer with mutation

operator for global optimization and application in training wavelet neural network. Journal of Intelligent & Fuzzy Systems, (Preprint), pp.1-14.
[3] Nibedan Panda Sir: Panda, N. and Majhi, S.K., 2020. Improved salp swarm algorithm with space transformation search for training neural network. Arabian Journal for Science and Engineering, 45(4), pp.2743-2761.
[4] Nibedan Panda Sir: Panda, N. and Majhi, S.K., Oppositional salp swarm algorithm with mutation operator for global optimization and application in training higher order neural networks. Multimedia Tools and Applications, pp.1-25.
[5] Nibedan Panda Sir: Panda, N. and Majhi, S.K., 2019. How effective is spotted hyena optimizer for training multilayer perceptrons. Int. J. Recent Technol. Eng, pp.4915-4927.
[6] Wong, R.J., Tran, T., Kaufman, H., Niles, J. and Gish, R., 2020. Geographic regions with high prevalence of nonalcoholic steatohepatitis-related hepatic fibrosis are also observed to demonstrate high prevalence of metabolic disease risk factors and low consumption of fruits and vegetables. Clinical Nutrition Experimental.
[7] Xia, M., Sun, X., Zheng, L., Bi, Y., Li, Q., Sun, L., Di, F., Li, H., Zhu, D., Gao, Y. and Bao, Y., 2020. Regional difference in the susceptibility of non-alcoholic fatty liver disease in China. BMJ Open Diabetes Research and Care, 8(1), p.e001311.
[8] Cannon, R.M., Davis, E.G., Goldberg, D.S., Lynch, R.J., Shah, M.B., Locke, J.E., McMasters, K.M. and Jones, C.M., 2020. Regional Variation in Appropriateness of Non-Hepatocellular Carcinoma Model for End-Stage Liver Disease Exception. Journal of the American College of Surgeons.
[9] Jothimani, D., Venugopal, R., Abedin, M.F., Kaliamoorthy, I. and Rela, M., 2020. COVID-19 and Liver. Journal of hepatology.
[10] Liao, H.Y., Hsu, J.C., Shia, B.C. and Lu, C.Y., 2020. PIN147 Geography Variations in Antibiotic Use RATE: Evidence from 8,026 Primary Clinics in Taiwan. Value in Health, 23, p.S568.
[11] Khorraminezhad, L., Leclercq, M., Droit, A., Bilodeau, J.F. and Rudkowska, I., 2020. Statistical and Machine-Learning Analyses in Nutritional Genomics Studies. Nutrients, 12(10), p.3140.
[12] Bendi, V.R. and Boddu, R.S.K., 2020. Performance Comparison of Classification Algorithms on Medical Datasets (No. 2322). EasyChair.
[13] Percival, B., Gibson, M., Leenders, J., Wilson, P.B. and Grootveld, M., 2020. Univariate and Multivariate Statistical Approaches to the Analysis and Interpretation of NMR-based Metabolomics Datasets of Increasing Complexity.
[14] Yousaf, M., Zahir, S., Riaz, M., Hussain, S.M. and Shah, K., 2020. Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. Chaos, Solitons & Fractals, 138, p.109926.
[15] Ustaoğlu, F., Tepe, Y. and Taş, B., 2020. Assessment of stream quality and health risk in a subtropical Turkey river system: A combined approach using statistical analysis and water quality index. Ecological Indicators, 113, p.105815.
[16] Petersen, M.W., Meyhoff, T.S., Helleberg, M., Kjær, M.B.N., Granholm, A., Hjortsø, C.J.S., Jensen, T.S., Møller, M.H., Hjortrup, P.B., Wetterslev, M. and Vesterlund, G.K., 2020. Low-dose hydrocortisone in patients with COVID-19 and severe hypoxia (COVID STEROID) trial—Protocol and statistical analysis plan. Acta Anaesthesiologica Scandinavica, 64(9), pp.1365-1375.
[17] Gyimah, R.A.A., Gyamfi, C., Anornu, G.K., Karikari, A.Y. and Tsyawo, F.W., 2020. Multivariate statistical analysis of water quality of the Densu River, Ghana. International Journal of River Basin Management, pp.1-11.
[18] Kattan, M.W. and Vickers, A.J., 2020. Statistical analysis and reporting guidelines for CHEST. Chest, 158(1), pp.S3-S11.
[19] Ren, R., Zhou, H., Hu, Z., He, S. and Wang, X., 2019. Statistical analysis of fire accidents in Chinese highway tunnels 2000–2016. Tunnelling and Underground Space Technology, 83, pp.452-460.
[20] Livadiotis, G., 2020. Statistical analysis of the impact of environmental temperature on the exponential growth rate of cases infected by COVID-19. PLoS one, 15(5), p.e0233875.
[21] Verma, A. and Ranga, V., 2018. Statistical analysis of CIDDS-001 dataset for network

intrusion detection systems using distance-based machine learning. Procedia Computer Science, 125, pp.709-716.

[22] Zajkowska, M., Gacuta, E., Kozłowska, S., Lubowicka, E., Głażewska, E.K., Chrostek, L., Szmitkowski, M., Pawłowski, P., Zbucka-Krętowska, M. and Ławicki, S., 2019. Diagnostic power of VEGF, MMP-9 and TIMP-1 in patients with breast cancer. A multivariate statistical analysis with ROC curve. Advances in medical sciences, 64(1), pp.1-8.

[23] Mustaqeem, A., Anwar, S.M., Khan, A.R. and Majid, M., 2017. A statistical analysis based recommender model for heart disease patients. International journal of medical informatics, 108, pp.134-145.

[24] Zhao, S., Gao, D., Zhuang, Z., Chong, M.K., Cai, Y., Ran, J., Cao, P., Wang, K., Lou, Y., Wang, W. and Yang, L., 2020. Estimating the serial interval of the novel coronavirus disease (COVID-19): A statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020.

[25] Lara-Cabrera, R., Gonzalez-Pardo, A. and Camacho, D., 2019. Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in Twitter. Future Generation Computer Systems, 93, pp.971-978.

[26] Barbulescu, A. and Nazzal, Y., 2020. Statistical analysis of dust storms in the United Arab Emirates. Atmospheric Research, 231, p.104669.

[27] Sidorov, O.V., Kozub, L.V., Goferberg, A.V. and Osintseva, N.V., 2018. Organization and Carrying out the Educational Experiment and Statistical Analysis of Its Results in IHL. European Journal of Contemporary Education, 7(1), pp.177-189.

[28] Nordhausen, K. and Oja, H., 2018. Independent component analysis: A statistical perspective. Wiley Interdisciplinary Reviews: Computational Statistics, 10(5), p.e1440.