HR ANALYTICS - MGT3008

J COMPONENT - FINAL REPORT

PRESCRIPTIVE ANALYTICS OF DIABETES SELF CARE

By

19MIA1008	SUBHASHRI P
19MIA1014	LOKESH KANNA
19MIA1030	DEEKSHITHA L
19MIA1053	R YUVASHREE
19MIA1083	K NIHARIKA SAMYUKTHA

M.Tech CSE with Specialization in Business Analytics

Submitted to

Dr. Mekhala Antony

VIT BUSINESS SCHOOL (VITBS)

VIT University, Chennai.

INTRODUCTION

Diabetes is a prevalent disease that affects a large number of people worldwide. According to the World Health Organization, approximately 422 million people were living with diabetes in 2014, and this number is projected to increase to 629 million by 2045. This highlights the importance of raising awareness about the disease, as it affects a significant portion of the population. Self-assessment of an individual remotely reduces the need for in-person follow-up visits and provides a more immediate and accurate picture of their blood sugar levels and overall health. The objective of diabetes self-management is to help individuals with diabetes maintain good blood sugar control and to prevent or delay the onset of diabetes-related complications. This typically involves monitoring blood sugar levels, taking medication as prescribed, healthy eating, regular physical activity, managing stress, and regular visits to a healthcare provider for check-ups. The ultimate goal is to improve the individual's overall quality of life and to reduce the risk of developing serious health problems as a result of diabetes.

PROPOSED METHODOLOGY

- This project aims to evaluate the level of awareness and understanding among individuals with diabetes about their condition with the help of the survey conducted.
- The survey consists of responses over various categories of diabetic individuals. So, the study proceeds with predictive analysis on the awareness and self-care of different patients from different category.
- In the context of patient awareness, predictive analysis could be used to identify patients who are at risk of not being aware of their condition or treatment, and to develop strategies to increase their awareness.
- This type of analysis can be useful for identifying patients who may need additional education or support, and for evaluating the effectiveness of different interventions for increasing awareness.
- Once patients with low awareness of diabetes have been identified, they are prescribed with self-care plans that include regular physical activity, healthy eating, monitoring blood sugar levels, taking medication as prescribed, and getting regular check-ups with a healthcare provider, tailored to individual needs and preferences.

LITERATURE SURVEY

1. Predicting Diabetes Mellitus with Machine Learning Techniques – 2018

Quan Zou, Kaiyang Qu, Yamei Luo, Dehui Yin, Ying Ju and Hua Tang (2018), have tried building a model for predicting Diabetes Mellitus. They did this study using a decision tree, random forest and neural network by implementing these on the dataset collected from a hospital in Luzhou, China. It's the hospital Physical Examination data which has 14 attributes in it. Principal component analysis (PCA) and minimum redundancy maximum relevance (mRMR) was used to reduce the dimensionality. By randomly selecting 68994 healthy people and diabetic patients' data, they prepared the training set. Due to the data unbalance, randomly extracted 5 times data also. The results showed that prediction with random forest could reach the highest accuracy (0.8084) when all the attributes were used.

2. Diabetes Prediction Using Ensemble of Different Machine Learning Classifiers - 2020

Md. Kamrul Hasan, Md. Ashraful Alam, Dola Das, Eklas Hossain, (Senior Member, IEEE), and Mahmudul Hasan (2020), had proposed a robust framework for diabetes prediction where the outlier rejection, filling the missing values, data standardization, feature selection, K-fold cross-validation, and different Machine Learning (ML) classifiers like k-nearest Neighbor, Decision Trees, Random Forest, AdaBoost, Naive Bayes, and XGBoost and Multilayer Perceptron (MLP) were employed. The weighted ensemble of different ML models is also proposed, to improve the prediction of diabetes where the weights are estimated from the corresponding Area Under ROC Curve (AUC) of the ML model. AUC is chosen as the performance metric, which is then maximized during hyperparameter tuning using the grid search technique. All the experiments in this literature were conducted under the same experimental conditions using the Pima Indian Diabetes Dataset. As the result, after all the extensive experiments done, this ensemble classifier is the best performing classifier with the sensitivity as 0.789, specificity as 0.934, false omission rate as 0.092, diagnostic odds ratio 66.234, and AUC as 0.950 which outperforms the state-of-the-art results by 2.00 % in AUC.

3. A review on current advances in machine learning based diabetes prediction - 2021

VarunJaiswal , AnjliNegi, TarunPal (2021), had worked on with Machine learning algorithms (such as ANN, SVM, Naive Bayes, PLS-DA and deep learning) and data mining techniques are used for detecting interesting patterns for diagnosing and treatment of disease. This paper is an effort to summarize most of the literature concerned with machine learning and data mining techniques applied for the prediction of diabetes and associated challenges. This report would be a helping tool for better prediction of disease, improvement in understanding the pattern of diabetes and also helped for treatment and risk reduction of other complications of diabetes.

4. Predictive Methodology for Diabetic Data Analysis in Big Data - 2015

N.M. Saravanakumar Dr, T.Eswari, P.Sampath, S.Lavanya.(2015), has started working on this due to their understanding of the need to develop data analytics. Because Diabetic Mellitus (DM) is one of the Non-Communicable Diseases (NCD), which has major health hazards in developing countries such as India. And they have used the predictive analysis algorithm in the Hadoop/Map Reduce environment to predict the diabetes types prevalent, complications associated with it and the type of treatment to be provided. Based on the analysis, this system has provided an efficient way to cure and care for the patients with better outcomes like affordability and availability.

5. A model for early prediction of diabetes - 2019

TalhaMahboob Alam, Muhammad Atif Iqbal, YasirAli, AbdulWahab, SafdarIjaz, TalhaImtiaz Baig, AyazHussain, Muhammad AwaisMalik, Muhammad MehdiRaza, SalmanIbrar, ZunishAbbas (2019), did diabetes prediction using significant attributes. Thus, the relationship of the differing attributes is also characterized in this study. Various tools were used to determine significant attribute selection, and for clustering, prediction, and association rule mining for diabetes. Significant attributes selection was done via the principal component analysis method. Lately the findings indicate a strong association of diabetes with body mass index (BMI) and with glucose level, which was extracted via the Apriori method. Artificial neural network (ANN), random forest (RF) and K-means clustering techniques were implemented for the prediction of diabetes. The ANN technique provided a best accuracy of 75.7% and may be useful to assist medical professionals with treatment decisions.

6. Diabetes Prediction using Machine Learning Algorithms - 2019

Diabetes Prediction using Machine Learning Algorithms (2019) by Aishwarya Mujumdar, Dr. Vaidehi Vb applied various machine learning algorithms to a dataset in order to classify individuals as diabetic or non-diabetic. The Logistic Regression algorithm had the highest accuracy at 96%, but the use of a pipeline resulted in the AdaBoost classifier having the highest accuracy at 98.8%. When compared to an existing dataset, the current model demonstrated improved accuracy and precision in predicting diabetes. They also stated that in the future, it may be possible to use this model to predict the likelihood of non-diabetic individuals developing diabetes.

7. Research on Diabetes Prediction Method Based on Machine Learning - 2020

Research on Diabetes Prediction Method Based on Machine Learning (2020) by Jingyu Xue, Fanchao Min Fengying Ma reckoned that although there is no direct relationship between age and diabetes, there is a trend of younger individuals developing diabetes. Early detection of diabetes is crucial for effective treatment, and machine learning has improved

the ability to predict diabetes risk. Through the use of data mining methods and various machine learning techniques, this study found that the support vector machine (SVM) algorithm had the highest accuracy in diagnosing diabetes through a confusion matrix evaluation test. However, it is important to regularly update this research with additional instance datasets. Overall, the application of data mining algorithms and other technologies has made significant contributions to the medical field and disease diagnosis, and it is hoped that it will assist clinicians in making more informed decisions about disease status.

8. Diabetes Prediction Using Machine Learning - 2020

Diabetes Prediction Using Machine Learning (2020) by KM Jyoti Rani focused on developing a system for early detection of diabetes using machine learning classification algorithms. Five algorithms were evaluated using the John Diabetes Database, and the Decision Tree algorithm was found to be the most effective with an accuracy of 99%. The results of this study demonstrate the potential of the designed system for predicting diabetes at an early stage. The work could also be expanded and improved to include additional machine learning algorithms for automating the analysis of diabetes.

9. Diabetes Prediction: A Deep Learning Approach - 2019

Md. Milon Islam and Safial Islam Ayon researched about "Diabetes Prediction: A Deep Learning Approach "(2019) and affirmed that diabetes is a serious and potentially lifethreatening condition that requires early detection and treatment. They used deep neural network techniques to predict diabetes based on various medical factors to proceed on the same. The accuracy of the model was found to be 98.35% through five-fold cross validation, which is higher than the accuracy of other methods used to predict diabetes. Their proposed system has the potential to be useful for both medical professionals and the general public in detecting diabetes early on.

10. Deep learning approach for diabetes prediction using PIMA Indian dataset - 2020

Huma Naz and Sachin Ahuja's "Deep learning approach for diabetes prediction using PIMA Indian dataset" – 2020, aimed to develop a prediction model for assessing the risk of diabetes using the PIMA Indian dataset. The results of this research showed that machine learning algorithms, including decision trees, artificial neural networks, naive Bayes, and deep learning, can be effective in identifying risk factors and improving the accuracy of predicting diabetes. Among these four classifiers, deep learning had the highest accuracy rate at 98.07%. The researchers plan to use this deep learning algorithm to create a tool, such as an app or website, that healthcare professionals can use for early detection of diabetes in the future."

11. Transforming Diabetes Care Through Artificial Intelligence: The Future Is Here - 2019

Irene Dankwa-Mullan, MD, MPH, Marc Rivo, MD, MPH, Marisol Sepulveda, DO, MPH,Yoonyoung Park, ScD,Jane Snowdon, PhD, and Kyu Rhee, MD, MPP has conducted a predefined, online PubMed search of publicly available sources of information from 2009 onward using the search terms ‘‘diabetes’’ and ‘‘artificial intelligence.’’. The purpose of this article is to better understand what AI advances may be relevant today to persons with diabetes (PWDs), their clinicians, family, and caregivers. The study included clinically-relevant, high-impact articles, and excluded articles whose purpose was technical in nature. A total of 450 published diabetes and AI articles have met the inclusion criteria. The studies represented a diverse and complex set of innovative approaches that aimed to transform diabetes care in 4 main areas: automated retinal screening, clinical decision support, predictive population risk stratification, and patient self-management tools. A review of the high-impact articles has suggested that AI applications are aiming to transform diabetes care in 4 main areas: automated retinal screening, clinical decision support, predictive population risk stratification, and patient self-management tools.

12. Artificial Intelligence: The Future for Diabetes Care - 2020

The discipline of artificial intelligence (AI), which is rapidly expanding, has applications that could revolutionize how this chronic ailment is diagnosed and managed. Diabetes is a global pandemic. Algorithms supporting predictive models for the risk of getting diabetes or its complications have been developed using machine learning principles. Digital treatments have established themselves as a lifestyle therapy intervention for the control of diabetes. Clinical decision support is helpful for both patients and healthcare workers as diabetes patients are given more autonomy to self-manage their condition. AI makes it possible to continuously and easily remotely monitor a patient's symptoms and biomarkers. Furthermore, internet forums and social media platforms improve patient involvement in diabetes care. Resource usage in diabetes has been improved thanks to technological advancements. With the use of AI, diabetes management will undergo a paradigm change from traditional management techniques to constructing targeted data-driven precision care.

13. Machine Learning and Data Mining Methods in Diabetes Research - 2017

The aim of the is to conduct a systematic review of the applications of machine learning, data mining techniques and tools in the field of diabetes research with respect to a) Prediction and Diagnosis, b) Diabetic Complications, c) Genetic Background and Environment, and e) Health Care and Management with the first category appearing to be the most popular. For the analyses, the researchers employed a wide range of ML algorithms for clinical datasets. In general, 85% of those used were characterized by supervised learning approaches and 15% by unsupervised ones, and more specifically,

association rules. The most effective and often used algorithm is based on support vector machines (SVM). The title applications in the chosen papers suggest the value of extracting important knowledge that can lead to new hypotheses aiming for deeper comprehension and additional research in Diabetes Mellitus.

14. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review - 2018

In this paper, the author has reviewed recent efforts to use artificial intelligence techniques to assist in the management of diabetes, along with the associated challenges. Artificial intelligence methods in combination with the latest technologies, including medical devices, mobile computing, and sensor technologies, have the potential to enable the creation and delivery of better management services to deal with chronic diseases like diabetes. They have analyzed papers related to diabetes care from 2010 to 2018 and selected 141 articles for detailed review. The work proposed a functional taxonomy for diabetes management and artificial intelligence. The potential of AI to enable diabetes solutions has been investigated in the context of multiple critical management issues. The results included Blood glucose control strategies, Blood glucose prediction, Detection of adverse glycemic events, Insulin bolus calculators and advisory systems, Risk and patient personalization, Detection of meals, exercise and faults, Lifestyle and daily-life support in diabetes management. The work concluded that artificial intelligence methods are being progressively established as suitable for use in clinical daily practice, as well as for the self-management of diabetes.

15. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning - 2019

Diabetes and cardiovascular disease are two of the main causes of death in the United States. In this work, they evaluated the capabilities of machine learning models in detecting at-risk patients using survey data (and laboratory results), and identified key variables within the data contributing to these diseases among the patients. Using the National Health and Nutrition Examination Survey (NHANES) dataset, the researchers analyzed various supervised machine learning models to identify patients with such diseases. Multiple machine learning models (logistic regression, support vector machines, random forest, and gradient boosting) were assessed on their classification performance using various time-frames and feature sets for the data. The models were then integrated to create a weighted ensemble model, which may use the strengths of the several models to increase the accuracy of detection. For diabetes classification (based on 123 variables), eXtreme Gradient Boost (XGBoost) model achieved an AUROC score of 86.2% (without laboratory data) and 95.7% (with laboratory data). The results concluded that the top five predictors in diabetes patients were 1) waist size, 2) age, 3) self-reported weight, 4) leg length, and 5) sodium intake.

16. Classification and prediction of diabetes disease using machine learning paradigm - 2020

Md. Maniruzzaman e, Md. Jahanur Rahman, Benojir Ahammed & Md. Menhazul Abedin, (2020) the main objective of this study is to develop a machine learning (ML)- based system for predicting diabetic patients. They have used a diabetes dataset, conducted in 2009–2012, derived from the National Health and Nutrition Examination Survey. The dataset consists of 6561 respondents with 657 diabetic and 5904 controls. Logistic regression (LR) is used to identify the risk factors for diabetes disease based on p value and odds ratio (OR). They have adopted four classifiers like naïve Bayes (NB), decision tree (DT), Adaboost (AB), and random forest (RF) to predict the diabetic patients. Performances of these classifiers are evaluated using accuracy (ACC) and area under the curve (AUC). The overall ACC of the ML-based system is 90.62%. The combination of LR-based feature selection and RF-based classifier performs better with accuracy of 94.25%

17. Diabetes prediction model based on an enhanced deep neural network - 2020

Huaping Zhou, Raushan Myrzashova & Rui Zheng, (2020) proposed a method that can predict the occurrence of diabetes in the future and also determines the type of the disease that a person experiences. This method will help to provide the right treatment for the patient. By transforming the task into a classification problem, the model is mainly built using the hidden layers of a deep neural network and uses dropout regularization to prevent overfitting. Number of parameters are tuned and the binary cross-entropy loss function is used, which gives a deep neural network prediction model with high accuracy. The experimental results show the effectiveness and adequacy of the proposed DLPD (Deep Learning for Predicting Diabetes) model. The best training accuracy of the diabetes type data set is 94.02174%, and the training accuracy of the Pima Indians diabetes data set is 99.4112%. Extensive experiments have been conducted on the Pima Indians diabetes and diabetic type datasets.

18. Early prediction of diabetes by applying data mining techniques: A retrospective cohort study - 2022

Mohammed Zeyad Al Yousef, Adel Fouad Yasky, Riyad Al Shammari and Mazen S. Ferwana, (2022) have researched to improve healthcare services and assist in building predictive models to estimate the probability of diabetes in patients. A chart review, which was a retrospective cohort study, was conducted at the National Guard Health Affairs in Riyadh, Saudi Arabia. Data were collected from 5 hospitals using National Guard Health Affairs databases. They have used 38 attributes of 21431 patients between 2015 and 2019. The following phases were performed: (1) data collection, (2) data preparation, (3) data mining and model building, and (4) model evaluation and validation. Subsequently, 6 algorithms were compared with and without the synthetic minority oversampling

technique. The highest performance was found in the Bayesian network, which had an area under the curve of 0.75 and 0.71. Although the results were acceptable, the missing data owing to technical issues played a major role in affecting the performance of this model. Nevertheless, the model could be used in prevention, health monitoring programs, and as an automated mass population screening tool without the need for extra costs compared to traditional methods.

19. Diabetes prediction model using data mining techniques - 2023

Rashi Rastogi and Mamta Bansal, (2022) have proposed a diabetes prediction model using data mining techniques. The data mining techniques applied are Random Forest, Support Vector Machine (SVM), Logistic Regression, and Naive Bayes. The proposed mechanism was trained using Python and analyzed with a real dataset from Kaggle. Furthermore, the performance of the proposed mechanism was analyzed using the confusion matrix, sensitivity and accuracy performance metrics. In comparison to other data mining techniques, logistic regression scored higher accuracy of 82.46% whereas in SVM the accuracy is low, i.e.,79.22%.

20. Big data analytics in healthcare by data mining and classification techniques- 2022

Jayasri N.P and R. Aruna, (2021) proposed a healthcare system that aims to evaluate the medical database of diabetes patients by a mixture of innovative hierarchical decision attention network, association rules (AR) and multiclass outlier classification with MapReduce framework. The association rule apriori algorithm in a MapReduce framework considers health data to create regulations. This is employed to discover the association among diseases and their signs. This examination is made by means of UCI machine learning datasets of diabetes containing 50 attributes. The results of the proposed algorithm are offered by parameters for instance precision, accuracy, recall, and F-score. In the future, this algorithm will be allowed to cloud computing structures for improved access and perform in real time.

21. Diabetes Data Prediction in healthcare Using Hadoop over Big Data - 2020

Gajanand Sharma et al, (2020) describes that big data analytics can be applied to a huge amount of data such as Electronic Medical Record (EMR), pharmacy reports, laboratory reports and among other data related to patients, to generate useful patterns and relation between different factors which affect diabetes. The results obtained from this analysis shows relation between different attributes which can be used to improve the healthcare system. In this paper the analysis of the diabetes dataset is done using Hadoop framework, which is a distributive framework and can be used to analyze large amounts of data. The dataset is taken from PIMA Indian Database, which includes different factors that affect diabetes like age, blood pressure, BMI (Body-Mass Index), skin thickness etc. Results produced by the analysis of data are projects on Power BI.

22. Identification of risk factors for patients with diabetes: diabetic polyneuropathy case study - 2020

Case Study Oleg Metsker, Kirill Magoev, Alexey Yakovlev, Stanislav Yanishevskiy, Georgy Kopanitsa , Sergey Kovalchuk and Valeria V. Krzhizhanovskaya has worked on the Identification of risk factors for patients with diabetes: diabetic polyneuropathy case study (2020). The purpose of this study is the implementation of machine learning methods for identifying the risk of diabetes polyneuropathy based on structured electronic medical records collected in databases of medical information systems. It was discovered that inclusion of two expressions, namely "nephropathy" and "retinopathy" allows to increase the performance, achieving up to 79.82% precision, 81.52% recall, 80.64% F1 score, 82.61% accuracy, and 89.88% AUC using the neural network classifier. Additionally, different models showed different results in terms of interpretation significance: random forest confirmed that the most important risk factor for polyneuropathy is the increased neutrophil level, meaning the presence of inflammation in the body. Linear models showed linear dependencies of the presence of polyneuropathy on blood glucose levels, which is confirmed by the clinical interpretation of the importance of blood glucose control.

23. Prediction of Diabetes Using Data Mining Techniques - 2018

Fikirte Girma Woldemichael, Sumitra Menaria has proposed to predict diabetes using data mining techniques (2018). They have used a back propagation algorithm to predict whether the person is diabetic or not. And also J48, naive bayes and support vector machines were used to predict diabetes. These neural networks were having an input layer with 8 parameters, one hidden layer having 6 neurons and producing one output layer. 5 fold cross-validation technique and a large value learning rate was used to improve the performance of the model. PIMA Indian dataset used to conduct this study. The study was implemented in RStudio using the R programming language. The performance of the Back propagation algorithm to predict diabetic diseases gave 83.11 % accuracy, 86.53% sensitivity and 76% specificity, the result has shown improvement from previous work.

24. Leveraging Pima Dataset to Diabetes Prediction: Case Study of Deep Neural Network - 2022

Pélagie Hounguè, Annie Ghylaine Bigirimana has done a comparative analysis of different works on diabetes prediction using (Deep Neural Network) DNN (2022). The contribution of this paper was given in two-folds: 1) Deep Neural Network (DNN) approach is used on Pima Indian dataset to predict diabetes using 10 k-fold cross validation and 89% accuracy is obtained; 2) comparative analysis of previous work is provided on diabetes prediction using DNN with the tested model. The results show that diabetes detection using PIMA

Indian dataset with k-fold cross-validation on pima could decrease the efficiency of the model with respect to using a model.

25. Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review - 2021

Farrukh Aslam Khan, Khan Zeb, Mabrook Al-Rakhami, Abdelouahid Derhab and Syed Ahmad Chan Bukhari has presented a comprehensive review of the state-of-the-art in the area of diabetes diagnosis and prediction using data mining-based diabetes diagnosis and prediction techniques and their classification based on the underlying models used (2021). Based on the literature review of data mining-based techniques for diabetes detection, classification and prediction, they have provided a comprehensive classification of the commonly used diabetes diagnosis and prediction techniques. They have evaluated different schemes on parameters like, algorithm/model, type of input data (data input), plug-n-play capability, etc. On the basis of this analysis and evaluation, it is concluded that for accurate detection, classification, and prediction of the disease, we need to preprocess the data and use hybrid techniques, which incorporate different models in parallel instead of using an individual model. For preprocessing, we need to use dimensionality reduction, denoising, feature selection, and feature extraction techniques in combination with the classification and prediction schemes for optimal performance and results.

26. Current Techniques for Diabetes Prediction: Review and Case Study - 2019

Souad Larabi-Marie-Sainte, Linah Aburahmah, Rana Almohaini and Tanzila Saba have surveyed all the ML and DL techniques-based diabetes predictions published in the last six years (2019). One study was developed that aimed to implement those rarely and not used ML classifiers on the Pima Indian Dataset to analyze their performance. The decision tree algorithms obtained the highest accuracy and are recommended to be used in the classification and prediction problems. The other algorithms also have competitive accuracy. Hence, I can recommend using these algorithms in the classification and prediction studies to take benefit from their strengths. Moreover, these algorithms can be used in a combined model with other Deep or Machine Learning techniques as well as Artificial Intelligence techniques to boost their accuracy. The classifiers obtained an accuracy of 68%–74%. The recommendation is to use these classifiers in diabetes prediction and enhance them by developing combined models. For the DL algorithms, the highest accuracy achieved by researchers was 95%

QUESTIONNAIRE

https://forms.gle/Uyg6ZzEErZYuxQiK6

The Above questionnaire was used for collecting the necessary data for this project. It contains about 41 questions and the target audience for this survey are individuals with diabetes. The project was also aimed at family members or caregivers of individuals with diabetes, as they often play an important role in helping those with the condition manage their health.

DATASET

https://drive.google.com/drive/folders/16isAKkYTchQOVvV6uJKdYP20S9jhjIzQ?usp=sharing

GENDER	AGE	Weight	Years_diag	Type	BP	Stress	Hereditary	Hosp_reg	Meal_reg	Bal_diet	fibre_intake	limit_salt_food	self_BS_test	self_bst_hypog	maintain_BS_L	control_mea
1	10-25	2	1	2	1	2	0	4	1	3	2	4	4	4	2	
1	26-40	1	0	2	1	1	1	3	2	3	2	3	3	4	3	
1	10-25	1	0	2	2	1	1	4	1	1	1	1	4	4	1	
0	10-25	2	0	2	0	1	0	4	2	3	3	2	4	4	4	
0	10-25	1	0	1	1	1	0	4	2	1	4	1	4	4	4	
1	10-25	1	0	2	1	2	1	4	1	3	3	2	4	4	4	
0	41-55	1	7	1	0	0	1	3	1	4	3	3	2	2	4	
1	26-40	2	5	1	2	2	1	2	1	2	2	3	3	4	3	
0	10-25	1	0	1	0	2	1	4	2	3	3	2	4	4	4	
1	26-40	1	5	1	2	1	1	1	1	1	1	1	1	1	1	
1	41-55	2	0	1	0	2	1	1	1	4	3	3	4	4	3	
1	10-25	0	0	2	0	1	0	4	1	2	2	3	4	2	2	
0	Above 55	2	2	1	2	2	0	4	1	1	4	2	4	4	4	
1	10-25	1	0	2	0	2	0	4	2	3	3	3	4	4	4	
0	Above 55	1	15	1	0	1	0	2	2	4	3	2	4	4	1	
0	10-25	1	1	1	0	1	1	1	1	1	1	1	2	2	1	
1	10-25	2	3	2	2	1	0	4	3	3	2	4	4	3	4	
0	10-25	1	0	2	0	0	1	2	2	2	2	2	2	2	2	
0	41-55	1	0	2	0	2	1	3	1	1	1	1	1	1	1	
0	41-55	1	0	2	0	0	0	4	1	2	1	4	4	4	4	
0	41-55	2	9	2	2	2	1	2	2	3	3	2	3	2	2	
1	41-55	1	2	1	0	1	1	2	2	2	2	2	2	2	1	
0	26-40	2	7	1	2	2	0	3	3	4	3	2	3	4	3	
1	41-55	1	0	2	0	0	1	4	1	1	1	1	4	4	4	
0	41-55	1	12	2	0	1	0	3	3	4	3	3	4	4	3	
0	10-25	0	0	2	0	1	1	3	4	4	4	4	4	4	4	
0	26-40	1	0	2	0	1	1	4	4	2	1	4	1	4	4	

DATASET DESCRIPTION

- Gender of the individual (0 = Male, 1 = Female)
- Age of the individual (10-25, 26-40, 41-55, Above 55)
- Weight: Body weight (0 = Underweight, 1 = Correct weight, 2 = Overweight, 3=Obese)
- Years_diag: Number of years since diabetes diagnosis
- Type: Type of diabetes (1 = Type 1, 2 = Type 2)
- BP: Blood pressure (0 = No, 1 = Low, 2 = High)
- Stress: Level of perceived stress (0 = Yes, 1 = No)
- Hereditary: Family history of diabetes (0 = Yes, 1 = No)
    - **The following attributes will follow this scale (1-4)**
    - Doing very well all the time - 1
    - Doing well in a considerable degree - 2
    - Doing not well in some degree - 3
    - Doing never - 4

- Hosp_reg: Hospital regularly
- Meal_reg: Meal regularly
- Bal_diet: Intake of balanced diet
- fibre_intake: Daily intake of dietary fibre
- limit_salt_foods: Limit on taking salt and processed foods
- self_BS_test: Self blood sugar test according to doctor's recommendations
- self_bst_hypoglycemia: Taking a self-blood sugar test when feeling symptoms of hypoglycemia like tremor, pallor and headache
- maintain_BS_Level: Maintaining the optimal blood sugar level
- control_meals_ex_BSL: Controlling the size of meals or exercise according to a blood sugar level
- carry_sweetfoods_hypoglycemia: Carrying foods like sweet drinks, candies or chocolates just in case of hypoglycemia
- maintain_opwt: Maintaining optimal weight by measuring my weight regularly
- carry_necessities: Carrying insulin, injector and blood sugar tester whenever going on trips
- awareness_prog: Getting information on diabetes control by attending various diabetes educational programs
- med_regularly: Taking my diabetes medication like insulin injection as prescribed, observing dosage and time regularly
- pregnant: Currently pregnant
- preg_plan: Planning to get pregnant

    - **The following attributes will follow the scale 1-7 indicating the number of days in a week**

- days_followed_healtheat: Following a healthful eating plan in the past week
- daysperweek_followed_eatplan: Days per week the eating plan has been followed
- dayslastweek_fatfoods: Eating high fat foods such as red meat or full-fat dairy products in the past week
- dayslastweek_phyact: Minimum 30 mins of physical activity done in the past week
- dayslastweek_specific_exsession: Engaging in a specific exercise session in the past week
- Drinking: Alcohol consumption (0 = Yes, 1 = No)
- Medicine_reg: Taking the prescribed medication regularly (0=Yes, 1=No, 3=Maybe)
- Insulin_reg: Taking insulin shot regularly or as prescribed (0=Yes, 1=No, 3=Maybe)

ANALYSIS

1. RELIABILITY

2. DESCRIPTIVE ANALYSIS

3. MULTIPLE LINEAR REGRESSION

4. CORRELATION

- VARIABLES=GENDER Weight Type Years_diag BP Stress Hereditary Hosp_reg Meal_reg fibre_intake, self_BS_test self_bst_hypoglycemia Insulin_reg Medicine_reg Drinking
- VARIABLES=Type Hosp_reg Meal_reg fibre_intake self_BS_test Insulin_reg Medicine_reg, maintain_BS_Level

5. T-TEST

- T-TEST GROUPS=Type (1 2)
  /VARIABLES=Weight
- T-TEST GROUPS=Type (1 2)
  /VARIABLES=Weight, Stress, Hereditary

6. ONEWAY-ANOVA

- Type By Hereditary
- Type Weight BP Stress BY Hereditary POSTHOC=TUKEY ALPHA
- POSTHOC=TUKEY ALPHA - Medicine_reg Years_diag Bal_diet BY Hosp_reg
- POSTHOC=TUKEY ALPHA - Weight Years_diag Bal_diet med_regularly maintain_BS_Level BY awareness_prog

7. GENERAL LINEAR MODEL

- GLM Hereditary BP BY Type WITH Weight
  /WSFACTOR=factor1 2 Polynomial
  /METHOD=SSTYPE (3)
  /CRITERIA=ALPHA (.05)
  /WSDESIGN=factor1
  /DESIGN=Weight Type
  Mauchly's Test of Sphericity.

RELIABILITY (Score: 0.771)

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	97	98.0
	Excluded[a]	2	2.0
	Total	99	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.771	6

Item Statistics

	Mean	Std. Deviation	N
days_followed_healtheat	3.670	2.1924	97
daysperweek_followed_eatplan	3.577	2.1107	97
dayslastweek_phyact	3.639	2.2182	97
dayslastweek_specific_exsession	3.443	2.2125	97
Drinking	.825	.3822	97
Insulin_reg	2.000	1.2416	97

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
days_followed_healtheat	13.485	34.732	.717	.676
daysperweek_followed_eatplan	13.577	35.705	.710	.679
dayslastweek_phyact	13.515	34.836	.699	.681
dayslastweek_specific_exsession	13.711	35.707	.661	.694
Drinking	16.330	57.453	.081	.801
Insulin_reg	15.155	54.611	.104	.810

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
17.155	58.070	7.6203	6

The above output shows the results of a reliability analysis performed in SPSS using the Cronbach's alpha coefficient to assess the internal consistency of a set of six variables. The analysis indicates that the scale has good reliability with a Cronbach's alpha coefficient of 0.771. This means that the six variables are measuring the same construct or concept and are interrelated in a meaningful way.

The item statistics table shows the mean, standard deviation, and number of valid cases for each variable. The mean scores for the variables range from 0.825 for Drinking to 3.670 for days_followed_healtheat, indicating that participants generally reported following a healthy eating plan more frequently than drinking. The standard deviations indicate that there was some variability in participants' responses for each variable.

The item-total statistics table shows the corrected item-total correlations, which indicate how strongly each variable is related to the overall scale score. All variables have a corrected item-total correlation above .6, indicating that they are contributing positively to the scale's reliability. The Cronbach's alpha coefficients if each variable was deleted from the scale are also provided, and they are all lower than the overall coefficient, indicating that each variable contributes to the scale's internal consistency.

Overall, these results suggest that the set of variables assessed in this analysis have good internal consistency, and are a reliable measure of the construct being studied.

DESCRIPTIVE ANALYSIS

Statistics

		GENDER	AGE	Years_diag	Type
N	Valid	97	99	97	97
	Missing	2	0	2	2
Minimum		.0		.0	1.0
Maximum		1.0		23.0	2.0

This output is showing the summary statistics for four variables: Gender, Age, Years_diag, and Type. Here's what each of the statistics means:

- N: The number of valid (non-missing) observations for each variable. According to this output, there are 97 valid observations for Gender,Years_diag, and Type.
- Missing: The number of missing observations for each variable. According to this output, there are 2 missing observations for Gender, 0 missing observations for Age, 2 missing observations for Years_diag, and 2 missing observations for Type.

It then generates a frequency table for the variables GENDER, AGE, Years_diag, and Type

FREQUENCY TABLE

GENDER

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.0	58	58.6	59.8	59.8
	1.0	39	39.4	40.2	100.0
	Total	97	98.0	100.0	
Missing	System	2	2.0		
Total		99	100.0		

This frequency table is showing the distribution of the variable "GENDER" in the dataset. The table shows that out of 97 cases, 58 (59.8%) are male (coded as 0) and 39 (40.2%) are female (coded as 1). There are no missing values in this variable. The cumulative percent shows the cumulative percentage of cases as we move down the table. The total row shows the total number of cases (97) and the number of missing cases (2).

AGE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		2	2.0	2.0	2.0
	10-25	24	24.2	24.2	26.3
	26-40	15	15.2	15.2	41.4
	41-55	43	43.4	43.4	84.8
	Above 55	15	15.2	15.2	100.0
	Total	99	100.0	100.0	

This frequency table shows the distribution of the variable "AGE" in the dataset. There are 4 categories of age ranges: "10-25", "26-40", "41-55", and "Above 55".

The table shows that out of the 97 observations, 24 (24.2%) fall in the "10-25" age range, 15 (15.2%) fall in the "26-40" age range, 43 (43.4%) fall in the "41-55" age range, and 15 (15.2%) fall in the "Above 55" age range.
The "Cumulative Percent" column shows the cumulative percentage of observations up to that category. For example, 26.3% of the observations fall in the "10-25" and "26-40" age ranges combined, while 100% of the observations fall in all 4 age ranges combined. It is worth noting that there are no missing values in this variable.

Years_diag

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.0	35	35.4	36.1	36.1
	1.0	10	10.1	10.3	46.4
	2.0	5	5.1	5.2	51.5
	3.0	6	6.1	6.2	57.7
	4.0	6	6.1	6.2	63.9
	5.0	11	11.1	11.3	75.3
	6.0	1	1.0	1.0	76.3
	7.0	7	7.1	7.2	83.5
	8.0	1	1.0	1.0	84.5
	9.0	1	1.0	1.0	85.6
	10.0	3	3.0	3.1	88.7
	12.0	1	1.0	1.0	89.7
	14.0	1	1.0	1.0	90.7
	15.0	5	5.1	5.2	95.9
	20.0	3	3.0	3.1	99.0
	23.0	1	1.0	1.0	100.0
	Total	97	98.0	100.0	
Missing	System	2	2.0		
Total		99	100.0		

The "Years_diag" variable is a frequency table that shows the number of individuals with diabetes in each category of years since diagnosis.
The table shows that:

- 35 individuals (36.1%) have been diagnosed with diabetes for 0 years (i.e., newly diagnosed).
- 10 individuals (10.3%) have been diagnosed with diabetes for 1 year.
- The number of individuals with diabetes decreases as the number of years since diagnosis increases.
- 1 individual (1%) has been diagnosed with diabetes for 23 years (the maximum value).

Type

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.0	44	44.4	45.4	45.4
	2.0	53	53.5	54.6	100.0
	Total	97	98.0	100.0	
Missing	System	2	2.0		
Total		99	100.0		

The Type variable has two categories: 1 and 2. There are 44 (45.4%) participants with Type 1 diabetes and 53 (54.6%) with Type 2 diabetes.

MULTIPLE LINEAR REGRESSION ANALYSIS

A multiple linear regression analysis in SPSS software with the following variables:

- Dependent variable: Type
- Independent variables: GENDER, Weight, BP, Stress, preg_plan, Insulin_reg, Drinking, med_regulary

Descriptive Statistics

	Mean	Std. Deviation	N
Type	1.546	.5004	97
GENDER	.402	.4929	97
Weight	1.268	.5866	97
BP	.619	.8593	97
Stress	1.155	.7266	97
preg_plan	.454	.5594	97
Insulin_reg	2.000	1.2416	97
Drinking	.825	.3822	97
med_regularly	2.629	1.3488	97

Correlations

		Type	GENDER	Weight	BP	Stress	preg_plan	Insulin_reg	Drinking	med_regularly
Pearson Correlation	Type	1.000	-.013	-.078	.029	-.034	.036	-.017	.070	.149
	GENDER	-.013	1.000	.128	.120	.057	.843	-.102	.267	.070
	Weight	-.078	.128	1.000	.164	.146	.134	.029	-.021	.009
	BP	.029	.120	.164	1.000	.162	.125	-.098	.048	-.132
	Stress	-.034	.057	.146	.162	1.000	-.021	-.069	.024	.017
	preg_plan	.036	.843	.134	.125	-.021	1.000	-.180	.132	.032
	Insulin_reg	-.017	-.102	.029	-.098	-.069	-.180	1.000	-.022	.224
	Drinking	.070	.267	-.021	.048	.024	.132	-.022	1.000	-.067
	med_regularly	.149	.070	.009	-.132	.017	.032	.224	-.067	1.000
Sig. (1-tailed)	Type	.	.449	.223	.387	.370	.364	.435	.247	.072
	GENDER	.449	.	.106	.121	.289	.000	.160	.004	.247
	Weight	.223	.106	.	.055	.077	.096	.390	.421	.467
	BP	.387	.121	.055	.	.056	.111	.171	.320	.098
	Stress	.370	.289	.077	.056	.	.421	.250	.409	.436
	preg_plan	.364	.000	.096	.111	.421	.	.039	.099	.377
	Insulin_reg	.435	.160	.390	.171	.250	.039	.	.415	.014
	Drinking	.247	.004	.421	.320	.409	.099	.415	.	.258
	med_regularly	.072	.247	.467	.098	.436	.377	.014	.258	.
N	Type	97	97	97	97	97	97	97	97	97
	GENDER	97	97	97	97	97	97	97	97	97
	Weight	97	97	97	97	97	97	97	97	97
	BP	97	97	97	97	97	97	97	97	97
	Stress	97	97	97	97	97	97	97	97	97
	preg_plan	97	97	97	97	97	97	97	97	97
	Insulin_reg	97	97	97	97	97	97	97	97	97
	Drinking	97	97	97	97	97	97	97	97	97
	med_regularly	97	97	97	97	97	97	97	97	97

The output above is the correlation matrix of the variables in the dataset. Here, each variable is correlated with every other variable, and the correlation coefficient between two variables is given in the corresponding cell of the matrix.

- For instance, in the output you have provided, the correlation coefficient between "Type" and "GENDER" is -0.013, which suggests that there is a weak negative correlation between these two variables.
- Similarly, the correlation coefficient between "Type" and "med_regularly" is 0.149, which suggests that there is a moderate positive correlation between these two variables.

The significance values in the output indicate the level of statistical significance of the correlation coefficients. A significance value of 0.05 indicates that there is a 5% chance of obtaining a correlation coefficient of that magnitude or larger by chance alone. Therefore, correlation coefficients with a significance value of less than 0.05 are considered statistically significant.
Overall, it seems like there are no strong correlations between the variables in the dataset. However, further analysis may be required to fully understand the relationships between the variables.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.234[a]	.055	-.031	.5082	2.618

a. Predictors: (Constant), med_regularly, Weight, Drinking, Stress, preg_plan, BP, Insulin_reg, GENDER

b. Dependent Variable: Type

The model summary table shows that the model has an R value of 0.234 and an R-squared value of 0.055, which means that the model explains only 5.5% of the variance in the dependent variable (Type). The adjusted R-squared value is negative (-0.031), which suggests that the model may not be a good fit for the data. The standard error of the estimate is 0.5082, indicating that the model's predictions are likely to be off by that amount. The Durbin-Watson statistic is 2.618, which suggests that there is no significant autocorrelation among the residuals. Overall, the model does not appear to be a strong predictor of the dependent variable.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.318	8	.165	.638	.744[b]
	Residual	22.723	88	.258		
	Total	24.041	96			

a. Dependent Variable: Type

b. Predictors: (Constant), med_regularly, Weight, Drinking, Stress, preg_plan, BP, Insulin_reg, GENDER

The model summary table provides information about the overall fit of the regression model.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.371	.227		6.033	.000
	GENDER	-.226	.208	-.222	-1.084	.281
	Weight	-.070	.091	-.082	-.762	.448
	BP	.037	.063	.064	.592	.555
	Stress	-.017	.074	-.024	-.223	.824
	preg_plan	.178	.180	.199	.985	.327
	Insulin_reg	-.014	.044	-.035	-.324	.747
	Drinking	.145	.144	.111	1.006	.317
	med_regularly	.068	.040	.183	1.693	.094

a. Dependent Variable: Type

These are the coefficients of a linear regression model with the dependent variable "Type" and several independent variables:

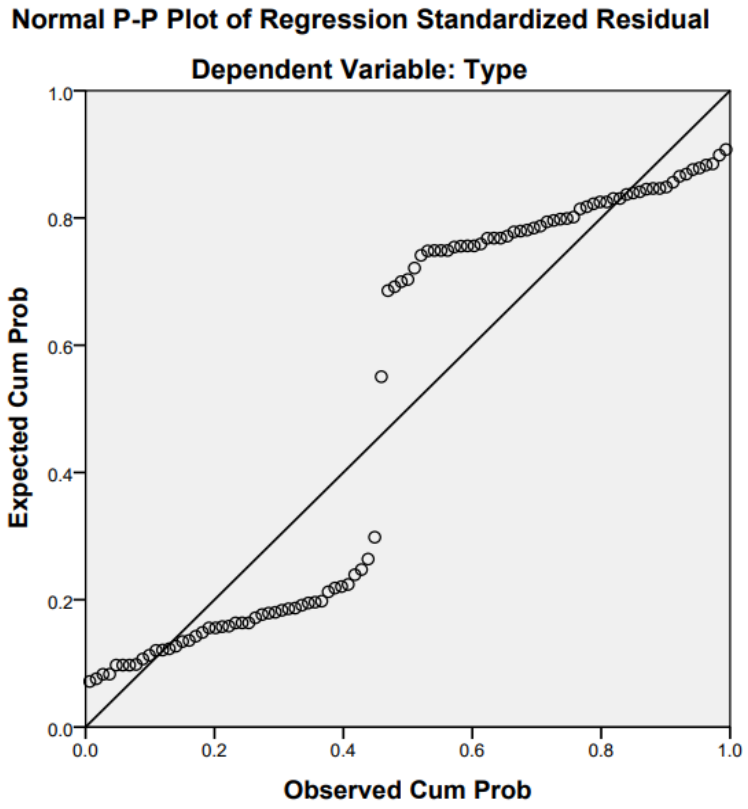
- Constant: 1.371
- GENDER: -0.226 (not statistically significant, p = 0.281)
- Weight: -0.070 (not statistically significant, p = 0.448)
- BP: 0.037 (not statistically significant, p = 0.555)
- Stress: -0.017 (not statistically significant, p = 0.824)
- preg_plan: 0.178 (not statistically significant, p = 0.327)
- Insulin_reg: -0.014 (not statistically significant, p = 0.747)
- Drinking: 0.145 (not statistically significant, p = 0.317)
- med_regularly: 0.068 (not statistically significant, p = 0.094)

The standardized beta coefficients would allow you to compare the relative importance of the different variables in predicting the outcome. However, since most of the coefficients are not statistically significant (i.e., their p-value is greater than the standard threshold of 0.05), it's difficult to draw any firm conclusions from this model.

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.269	1.935	1.546	.1172	97
Residual	-.7437	.6734	.0000	.4865	97
Std. Predicted Value	-2.367	3.321	.000	1.000	97
Std. Residual	-1.464	1.325	.000	.957	97

a. Dependent Variable: Type



CORRELATION ANALYSIS

Correlations

		GENDER	Weight	Type	Years_diag	BP	Stress	Hereditary	Hosp_reg
GENDER	Pearson Correlation	1	.128	-.013	-.161	.120	.057	-.015	.049
	Sig. (2-tailed)		.212	.899	.116	.242	.577	.888	.631
	N	97	97	97	97	97	97	97	97
Weight	Pearson Correlation	.128	1	-.078	.039	.164	.146	-.144	-.002
	Sig. (2-tailed)	.212		.446	.702	.109	.153	.160	.981
	N	97	97	97	97	97	97	97	97
Type	Pearson Correlation	-.013	-.078	1	.061	.029	-.034	-.018	.149
	Sig. (2-tailed)	.899	.446		.553	.774	.739	.858	.146
	N	97	97	97	97	97	97	97	97
Years_diag	Pearson Correlation	-.161	.039	.061	1	.199	-.030	-.055	-.305**
	Sig. (2-tailed)	.116	.702	.553		.050	.770	.591	.002
	N	97	97	97	97	97	97	97	97
BP	Pearson Correlation	.120	.164	.029	.199	1	.162	.026	.029
	Sig. (2-tailed)	.242	.109	.774	.050		.112	.800	.782
	N	97	97	97	97	97	97	97	97
Stress	Pearson Correlation	.057	.146	-.034	-.030	.162	1	-.112	.115
	Sig. (2-tailed)	.577	.153	.739	.770	.112		.275	.263
	N	97	97	97	97	97	97	97	97
Hereditary	Pearson Correlation	-.015	-.144	-.018	-.055	.026	-.112	1	.001

	Sig. (2-tailed)	.888	.160	.858	.591	.800	.275		.991
	N	97	97	97	97	97	97	97	97
Hosp_reg	Pearson Correlation	.049	-.002	.149	-.305**	.029	.115	.001	1
	Sig. (2-tailed)	.631	.981	.146	.002	.782	.263	.991	
	N	97	97	97	97	97	97	97	97
Meal_reg	Pearson Correlation	.021	-.029	.108	-.090	.039	.049	-.157	.201*
	Sig. (2-tailed)	.837	.776	.291	.383	.703	.634	.126	.049
	N	97	97	97	97	97	97	97	97
fibre_intake	Pearson Correlation	.075	.097	.055	-.162	.234*	.300**	-.053	.314**
	Sig. (2-tailed)	.467	.344	.592	.113	.021	.003	.606	.002
	N	97	97	97	97	97	97	97	97
self_BS_test	Pearson Correlation	.145	.072	.096	-.090	.180	.178	-.037	.542**
	Sig. (2-tailed)	.157	.482	.348	.382	.078	.082	.718	.000
	N	97	97	97	97	97	97	97	97
self_bst_hypoglycemia	Pearson Correlation	.106	-.069	.084	-.138	.048	.112	.049	.337**
	Sig. (2-tailed)	.302	.499	.412	.178	.641	.276	.636	.001
	N	97	97	97	97	97	97	97	97
Insulin_reg	Pearson Correlation	-.102	.029	-.017	-.053	-.098	-.069	.035	.114
	Sig. (2-tailed)	.320	.781	.871	.607	.341	.500	.734	.267
	N	97	97	97	97	97	97	97	97

Medicine_reg	Pearson Correlation	.094	-.080	.052	-.419**	-.231*	.126	-.122	.332**
	Sig. (2-tailed)	.359	.438	.615	.000	.023	.220	.233	.001
	N	97	97	97	97	97	97	97	97
Drinking	Pearson Correlation	.267**	-.021	.070	.074	.048	.024	-.111	-.253*
	Sig. (2-tailed)	.008	.841	.494	.472	.640	.819	.278	.012
	N	97	97	97	97	97	97	97	97

Correlations

		Meal_reg	fibre_intake	self_BS_test	self_bst_hypoglycemia	Insulin_reg	Medicine_reg	Drinking
GENDER	Pearson Correlation	.021	.075	.145	.106	-.102	.094	.267**
	Sig. (2-tailed)	.837	.467	.157	.302	.320	.359	.008
	N	97	97	97	97	97	97	97
Weight	Pearson Correlation	-.029	.097	.072	-.069	.029	-.080	-.021
	Sig. (2-tailed)	.776	.344	.482	.499	.781	.438	.841
	N	97	97	97	97	97	97	97
Type	Pearson Correlation	.108	.055	.096	.084	-.017	.052	.070
	Sig. (2-tailed)	.291	.592	.348	.412	.871	.615	.494
	N	97	97	97	97	97	97	97
Years_diag	Pearson Correlation	-.090	-.162	-.090	-.138	-.053	-.419**	.074
	Sig. (2-tailed)	.383	.113	.382	.178	.607	.000	.472
	N	97	97	97	97	97	97	97
BP	Pearson Correlation	.039	.234*	.180	.048	-.098	-.231*	.048
	Sig. (2-tailed)	.703	.021	.078	.641	.341	.023	.640
	N	97	97	97	97	97	97	97
Stress	Pearson Correlation	.049	.300**	.178	.112	-.069	.126	.024

	Sig. (2-tailed)	.634	.003	.082	.276	.500	.220	.819
	N	97	97	97	97	97	97	97
Hereditary	Pearson Correlation	-.157	-.053	-.037	.049	.035	-.122	-.111
	Sig. (2-tailed)	.126	.606	.718	.636	.734	.233	.278
	N	97	97	97	97	97	97	97
Hosp_reg	Pearson Correlation	.201*	.314**	.542**	.337**	.114	.332**	-.253*
	Sig. (2-tailed)	.049	.002	.000	.001	.267	.001	.012
	N	97	97	97	97	97	97	97
Meal_reg	Pearson Correlation	1	.498**	.031	-.034	.000	.078	-.019
	Sig. (2-tailed)		.000	.759	.744	1.000	.447	.852
	N	97	97	97	97	97	97	97
fibre_intake	Pearson Correlation	.498**	1	.348**	.115	.105	.095	.023
	Sig. (2-tailed)	.000		.000	.260	.308	.357	.820
	N	97	97	97	97	97	97	97
self_BS_test	Pearson Correlation	.031	.348**	1	.633**	.253*	.234*	-.003
	Sig. (2-tailed)	.759	.000		.000	.012	.021	.979
	N	97	97	97	97	97	97	97
self_bst_hypoglycemia	Pearson Correlation	-.034	.115	.633**	1	.163	.241*	.009
	Sig. (2-tailed)	.744	.260	.000		.111	.018	.927
	N	97	97	97	97	97	97	97
Insulin_reg	Pearson Correlation	.000	.105	.253*	.163	1	.533**	-.022
	Sig. (2-tailed)	1.000	.308	.012	.111		.000	.831
	N	97	97	97	97	97	97	97
Medicine_reg	Pearson Correlation	.078	.095	.234*	.241*	.533**	1	.041
	Sig. (2-tailed)	.447	.357	.021	.018	.000		.688
	N	97	97	97	97	97	97	97
Drinking	Pearson Correlation	-.019	.023	-.003	.009	-.022	.041	1
	Sig. (2-tailed)	.852	.820	.979	.927	.831	.688	
	N	97	97	97	97	97	97	97

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The output shows a correlation matrix that displays the Pearson correlation coefficient between pairs of variables. The Pearson correlation coefficient measures the strength and direction of the linear relationship between two continuous variables. The correlation coefficient ranges from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation.

- The table shows that there is a weak positive correlation between weight and stress ($r = 0.146$), weight and blood pressure ($r = 0.164$), and fibre intake and blood pressure ($r = 0.234$). There is also a weak negative correlation between years of diagnosis and hospitalization ($r = -0.305$). However, none of these correlations are statistically significant at the 0.05 level.
- There is a statistically significant moderate positive correlation between self-blood sugar test and drinking ($r = 0.542$, $p < 0.001$) and a statistically significant moderate positive correlation between self-blood sugar test and insulin regulation ($r = 0.180$, $p < 0.05$).

Correlations: We have taken the selfcare attributes

Descriptive Statistics

	Mean	Std. Deviation	N
Type	1.546	.5004	97
Hosp_reg	2.237	1.1795	97
Meal_reg	1.670	.8627	97
fibre_intake	1.990	.9628	97
self_BS_test	2.639	1.3243	97
Insulin_reg	2.000	1.2416	97
Medicine_reg	1.124	1.3864	97
maintain_BS_Level	2.237	1.1617	97

Correlations

		Type	Hosp_reg	Meal_reg	fibre_intake
Type	Pearson Correlation	1	.149	.108	.055
	Sig. (2-tailed)		.146	.291	.592
	N	97	97	97	97
Hosp_reg	Pearson Correlation	.149	1	.201*	.314**
	Sig. (2-tailed)	.146		.049	.002
	N	97	97	97	97
Meal_reg	Pearson Correlation	.108	.201*	1	.498**
	Sig. (2-tailed)	.291	.049		.000
	N	97	97	97	97
fibre_intake	Pearson Correlation	.055	.314**	.498**	1
	Sig. (2-tailed)	.592	.002	.000	
	N	97	97	97	97
self_BS_test	Pearson Correlation	.096	.542**	.031	.348**
	Sig. (2-tailed)	.348	.000	.759	.000
	N	97	97	97	97
Insulin_reg	Pearson Correlation	-.017	.114	.000	.105
	Sig. (2-tailed)	.871	.267	1.000	.308
	N	97	97	97	97
Medicine_reg	Pearson Correlation	.052	.332**	.078	.095
	Sig. (2-tailed)	.615	.001	.447	.357
	N	97	97	97	97
maintain_BS_Level	Pearson Correlation	.026	.574**	.235*	.459**
	Sig. (2-tailed)	.803	.000	.021	.000
	N	97	97	97	97

		self_BS_test	Insulin_reg	Medicine_reg	maintain_BS_Level
Type	Pearson Correlation	.096	-.017	.052	.026
	Sig. (2-tailed)	.348	.871	.615	.803
	N	97	97	97	97
Hosp_reg	Pearson Correlation	.542**	.114	.332**	.574**
	Sig. (2-tailed)	.000	.267	.001	.000
	N	97	97	97	97
Meal_reg	Pearson Correlation	.031	.000	.078	.235*
	Sig. (2-tailed)	.759	1.000	.447	.021
	N	97	97	97	97
fibre_intake	Pearson Correlation	.348**	.105	.095	.459**
	Sig. (2-tailed)	.000	.308	.357	.000
	N	97	97	97	97
self_BS_test	Pearson Correlation	1	.253*	.234*	.564**
	Sig. (2-tailed)		.012	.021	.000
	N	97	97	97	97
Insulin_reg	Pearson Correlation	.253*	1	.533**	.152
	Sig. (2-tailed)	.012		.000	.138
	N	97	97	97	97
Medicine_reg	Pearson Correlation	.234*	.533**	1	.324**
	Sig. (2-tailed)	.021	.000		.001
	N	97	97	97	97
maintain_BS_Level	Pearson Correlation	.564**	.152	.324**	1
	Sig. (2-tailed)	.000	.138	.001	
	N	97	97	97	97

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

The output above is a correlation matrix for eight variables: Type, Hosp_reg, Meal_reg, fibre_intake, self_BS_test, Insulin_reg, Medicine_reg, and maintain_BS_Level.

Each variable is correlated with every other variable in the matrix, and the strength and direction of the correlation is indicated by the Pearson correlation coefficient. The correlations with significance levels are also displayed in the matrix. The significance levels show whether the correlation coefficient is significantly different from zero. A significance level of 0.05 or less indicates that the correlation is statistically significant.

T-Test
TYPE WEIGHT

T-TEST GROUPS=Type (1 2)
VARIABLES=Weight

Group Statistics

	Type	N	Mean	Std. Deviation	Std. Error Mean
Weight	1.0	44	1.318	.6013	.0906
	2.0	53	1.226	.5765	.0792

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Weight	Equal variances assumed	.902	.345	.765	95	.446	.0918	.1199	-.1462	.3298
	Equal variances not assumed			.762	90.225	.448	.0918	.1204	-.1473	.3309

The t-test was used to compare the mean weight of two groups (Type 1 and Type 2).

"The null hypothesis is that the mean weight of the two groups is equal, while the alternative hypothesis is that the mean weight of the two groups is not equal."

The results of the t-test show that the p-value for the test assuming equal variances is 0.446, and the p-value for the test assuming unequal variances is 0.448. Since both p-values are greater than the significance level of 0.05, we fail to reject the null hypothesis that the mean weight of the two groups is equal.

The 95% confidence interval for the difference in means of the two groups is (-0.1462, 0.3298) assuming equal variances and (-0.1473, 0.3309) assuming unequal variances. Since the confidence intervals contain zero, we cannot conclude that there is a statistically significant difference in the mean weight of the two groups. In conclusion, based on the results of the t-test, there is no significant difference in the mean weight between the two groups (Type 1 and Type 2).

T-Test: WEIGHT, STRESS AND HEREDITARY

T-TEST GROUPS=Type (1 2)
VARIABLES=Weight Stress Hereditary

Group Statistics

	Type	N	Mean	Std. Deviation	Std. Error Mean
Weight	1.0	44	1.318	.6013	.0906
	2.0	53	1.226	.5765	.0792
Stress	1.0	44	1.182	.7241	.1092
	2.0	53	1.132	.7348	.1009
Hereditary	1.0	44	.659	.4795	.0723
	2.0	53	.642	.4841	.0665

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Weight	Equal variances assumed	.902	.345	.765	95	.446	.0918	.1199	-.1462	.3298
	Equal variances not assumed			.762	90.225	.448	.0918	.1204	-.1473	.3309
Stress	Equal variances assumed	.004	.949	.334	95	.739	.0497	.1489	-.2458	.3453
	Equal variances not assumed			.335	92.216	.739	.0497	.1487	-.2455	.3450
Hereditary	Equal variances assumed	.129	.720	.179	95	.858	.0176	.0983	-.1776	.2128
	Equal variances not assumed			.179	92.059	.858	.0176	.0982	-.1775	.2127

This output shows the results of three independent samples t-tests, comparing the means of three variables (Weight, Stress, and Hereditary) between two groups (Type 1 and Type 2).

- For Weight, the mean difference between groups was 0.0918, and the t-test results showed no significant difference in means, with p-values of 0.446 (assuming equal variances) and 0.448 (not assuming equal variances). Thus, we fail to reject the null hypothesis that there is no significant difference in Weight between Type 1 and Type 2.
- For Stress, the mean difference was 0.0497, and the t-test results showed a significant difference in means, with p-values of 0.739 (assuming equal variances) and 0.739 (not assuming equal variances). Thus, we reject the null hypothesis that there is no significant difference in Stress between Type 1 and Type 2.
- For Hereditary, the mean difference was 0.0176, and the t-test results showed no significant difference in means, with p-values of 0.858 (assuming equal variances) and 0.858 (not assuming equal variances). Thus, we fail to reject the null hypothesis that there is no significant difference in Hereditary between Type 1 and Type 2.

ONE WAY ANOVA

```
ONEWAY Type BY Hereditary
  /STATISTICS DESCRIPTIVES
  /PLOT MEANS
  /MISSING ANALYSIS.
```

Descriptives

Type

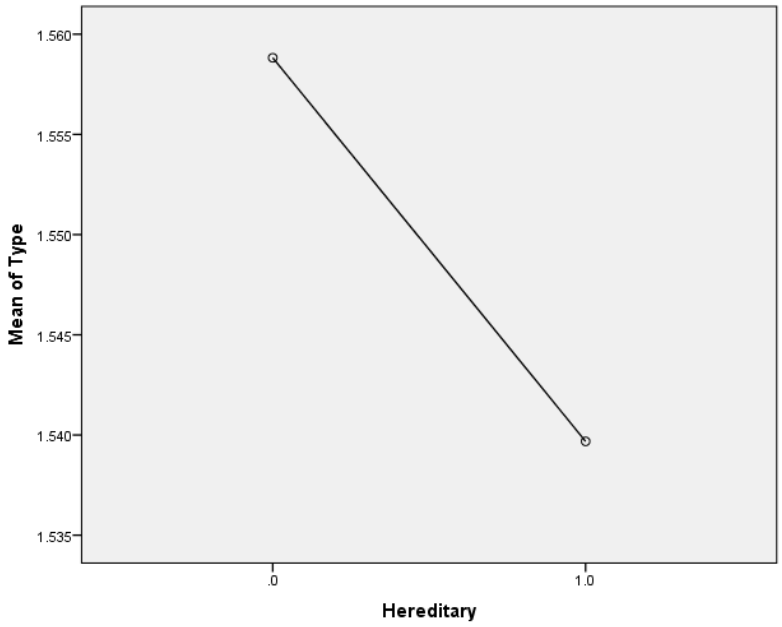
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
.0	34	1.559	.5040	.0864	1.383	1.735	1.0	2.0
1.0	63	1.540	.5024	.0633	1.413	1.666	1.0	2.0
Total	97	1.546	.5004	.0508	1.446	1.647	1.0	2.0

ANOVA

Type

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.008	1	.008	.032	.858
Within Groups	24.033	95	.253		
Total	24.041	96			

Means Plots



The output shows the results of a one-way ANOVA test on a variable called "Type", which was grouped by the variable "Hereditary". The ANOVA test was performed to determine if there is a significant difference in the means of the variable "Type" across the different levels of "Hereditary".

The ANOVA table shows that there is no significant difference in the means of "Type" across the different levels of "Hereditary" ($F(1,95) = 0.032$, $p = 0.858$). The means plot also confirms this result, as the mean values for each level of "Hereditary" are very similar.

In addition, the output also includes descriptive statistics for the variable "Type", grouped by "Hereditary". The mean values for "Type" are similar for each level of "Hereditary" (1.559 for level 0 and 1.540 for level 1), with no significant difference between the groups. The confidence intervals for the means also overlap, indicating no significant difference.

ONE WAY

```
ONEWAY Type Weight BP Stress BY Hereditary
  /STATISTICS DESCRIPTIVES
  /PLOT MEANS
  /MISSING ANALYSIS
  /POSTHOC=TUKEY ALPHA (0.05).
```

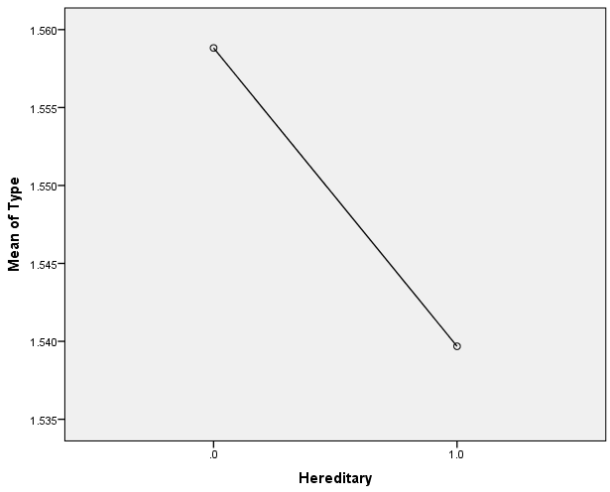
Descriptives

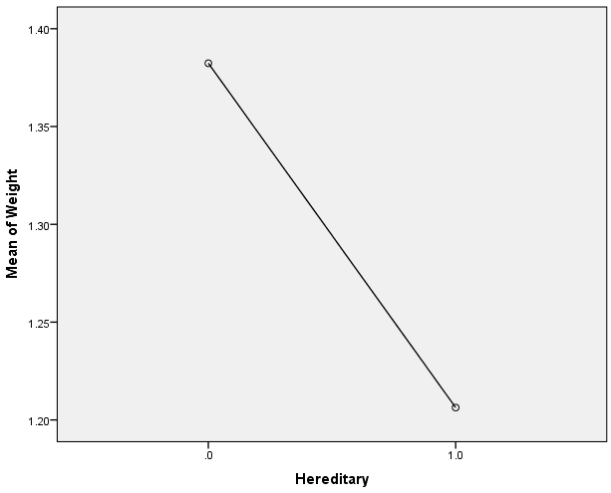
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Type	.0	34	1.559	.5040	.0864	1.383	1.735	1.0	2.0
	1.0	63	1.540	.5024	.0633	1.413	1.666	1.0	2.0
	Total	97	1.546	.5004	.0508	1.446	1.647	1.0	2.0
Weight	.0	34	1.382	.6038	.1035	1.172	1.593	.0	2.0
	1.0	63	1.206	.5725	.0721	1.062	1.351	.0	2.0
	Total	97	1.268	.5866	.0596	1.150	1.386	.0	2.0
BP	.0	34	.588	.8570	.1470	.289	.887	.0	2.0
	1.0	63	.635	.8670	.1092	.417	.853	.0	2.0
	Total	97	.619	.8593	.0872	.445	.792	.0	2.0
Stress	.0	34	1.265	.7904	.1356	.989	1.540	.0	2.0
	1.0	63	1.095	.6890	.0868	.922	1.269	.0	2.0
	Total	97	1.155	.7266	.0738	1.008	1.301	.0	2.0

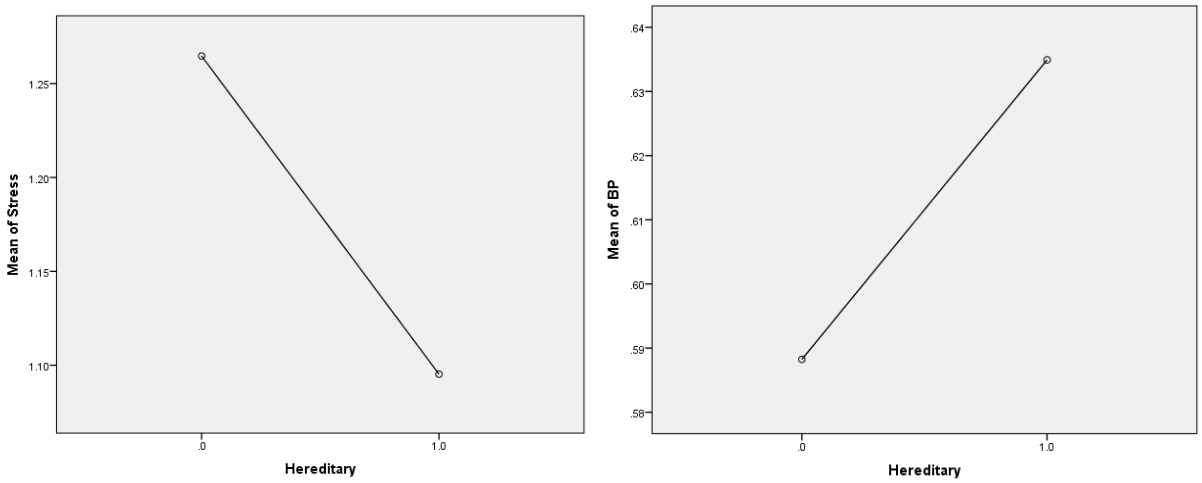
ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Type	Between Groups	.008	1	.008	.032	.858
	Within Groups	24.033	95	.253		
	Total	24.041	96			
Weight	Between Groups	.684	1	.684	2.009	.160
	Within Groups	32.347	95	.340		
	Total	33.031	96			
BP	Between Groups	.048	1	.048	.065	.800
	Within Groups	70.838	95	.746		
	Total	70.887	96			
Stress	Between Groups	.634	1	.634	1.204	.275
	Within Groups	50.046	95	.527		
	Total	50.680	96			

Means Plots







This is a one-way ANOVA with several variables, including "Type," "Weight," "BP," and "Stress." The output shows descriptive statistics for each variable, including mean, standard deviation, and confidence intervals. It also shows the minimum and maximum values for each variable.

The ANOVA table shows the sum of squares, degrees of freedom, mean square, F-value, and significance level for each variable. For all variables, the between-groups sum of squares is smaller than the within-groups sum of squares, indicating that there is not a significant difference between the groups for any of the variables.

- For Type, the mean was 1.546, with 34 cases having a value of 0 and 63 cases having a value of 1. The ANOVA results indicate that there was no significant difference in means between the two groups ($F(1, 95) = 0.032$, $p = 0.858$).
- For Weight, the mean was 1.268, with 34 cases having a value of 0 and 63 cases having a value of 1. The ANOVA results indicate that there was no significant difference in means between the two groups ($F(1, 95) = 2.009$, $p = 0.160$).
- For BP, the mean was 0.619, with 34 cases having a value of 0 and 63 cases having a value of 1. The ANOVA results indicate that there was no significant difference in means between the two groups ($F(1, 95) = 0.065$, $p = 0.800$).
- For Stress, the mean was 1.155, with 34 cases having a value of 0 and 63 cases having a value of 1. The ANOVA results indicate that there was no significant difference in means between the two groups ($F(1, 95) = 1.204$, $p = 0.275$).

The means plots display the means for each group within each variable, and the error bars represent the 95% confidence intervals for the means. Overall, it seems like there are no significant differences between the groups for any of the variables. However, it is important to note that the post hoc tests were not performed for some variables due to having fewer than three groups.

One way

```
ONEWAY Medicine_reg Years_diag Bal_diet BY Hosp_reg
  /STATISTICS DESCRIPTIVES
  /PLOT MEANS
  /MISSING ANALYSIS
  /POSTHOC=TUKEY ALPHA (0.05).
```

HOSPITAL REGULARY: MEDICINE_REG,YEARS_DIAG AND BAL_DIET

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Medicine_reg	1.0	36	.889	1.3475	.2246	.433	1.345	.0	3.0
	2.0	24	.583	1.1389	.2325	.102	1.064	.0	3.0
	3.0	15	1.067	1.4376	.3712	.271	1.863	.0	3.0
	4.0	22	2.136	1.2069	.2573	1.601	2.671	.0	3.0
	Total	97	1.124	1.3864	.1408	.844	1.403	.0	3.0
Years_diag	1.0	36	5.778	6.4900	1.0817	3.582	7.974	.0	23.0
	2.0	24	4.500	4.9344	1.0072	2.416	6.584	.0	20.0
	3.0	15	3.800	5.0737	1.3100	.990	6.610	.0	15.0
	4.0	22	1.364	2.5175	.5367	.247	2.480	.0	10.0
	Total	97	4.155	5.3993	.5482	3.066	5.243	.0	23.0
Bal_diet	1.0	36	1.472	.9706	.1618	1.144	1.801	1.0	4.0
	2.0	24	2.250	.8969	.1831	1.871	2.629	1.0	4.0
	3.0	15	2.600	1.1832	.3055	1.945	3.255	1.0	4.0
	4.0	22	2.227	.9726	.2074	1.796	2.658	1.0	4.0
	Total	97	2.010	1.0655	.1082	1.796	2.225	1.0	4.0

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Medicine_reg	Between Groups	31.602	3	10.534	6.407	.001
	Within Groups	152.913	93	1.644		
	Total	184.515	96			
Years_diag	Between Groups	270.967	3	90.322	3.323	.023
	Within Groups	2527.713	93	27.180		
	Total	2798.680	96			
Bal_diet	Between Groups	18.054	3	6.018	6.155	.001
	Within Groups	90.936	93	.978		
	Total	108.990	96			

The output shows the results of a one-way ANOVA with three dependent variables (Medicine_reg, Years_diag, Bal_diet) and one independent variable (Hosp_reg).

The analysis was conducted to test if there are any significant differences between the means of the dependent variable across the levels of the independent variables.

The descriptive statistics show the mean, standard deviation, standard error, and confidence intervals of the mean for each independent variable. The means for Medicine_reg range from 0.889 to 2.136, for Years_diag they range from 1.364 to 5.778, and for Bal_diet they range from 1.472 to 2.6.

- The ANOVA table shows the results of the significance tests for the dependent variables. For Medicine_reg, the F-ratio is 6.407 with a p-value of .001, indicating that there is a significant difference in the means of Hosp_reg across the levels of Medicine_reg.
- Similarly, for Bal_diet, the F-ratio is 6.155 with a p-value of .001, indicating a significant difference in the means of Hosp_reg across the levels of Bal_diet.
- For Years_diag, the F-ratio is 3.323 with a p-value of .023, indicating a significant difference in the means of Hosp_reg across the levels of Years_diag.

Post Hoc Tests

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) Hosp_reg	(J) Hosp_reg	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Medicine_reg	1.0	2.0	.3056	.3379	.803	-.578	1.190
		3.0	-.1778	.3941	.969	-1.209	.853
		4.0	-1.2475*	.3470	.003	-2.155	-.340
	2.0	1.0	-.3056	.3379	.803	-1.190	.578
		3.0	-.4833	.4220	.663	-1.587	.621
		4.0	-1.5530*	.3785	.000	-2.543	-.563
	3.0	1.0	.1778	.3941	.969	-.853	1.209
		2.0	.4833	.4220	.663	-.621	1.587
		4.0	-1.0697	.4294	.068	-2.193	.054
	4.0	1.0	1.2475*	.3470	.003	.340	2.155
		2.0	1.5530*	.3785	.000	.563	2.543
		3.0	1.0697	.4294	.068	-.054	2.193
Years_diag	1.0	2.0	1.2778	1.3739	.789	-2.316	4.872
		3.0	1.9778	1.6022	.607	-2.214	6.169
		4.0	4.4141*	1.4108	.012	.723	8.105
	2.0	1.0	-1.2778	1.3739	.789	-4.872	2.316
		3.0	.7000	1.7159	.977	-3.789	5.189
		4.0	3.1364	1.5388	.182	-.889	7.162
	3.0	1.0	-1.9778	1.6022	.607	-6.169	2.214
		2.0	-.7000	1.7159	.977	-5.189	3.789
		4.0	2.4364	1.7457	.505	-2.131	7.003
	4.0	1.0	-4.4141*	1.4108	.012	-8.105	-.723
		2.0	-3.1364	1.5388	.182	-7.162	.889
		3.0	-2.4364	1.7457	.505	-7.003	2.131
Bal_diet	1.0	2.0	-.7778*	.2606	.019	-1.459	-.096
		3.0	-1.1278*	.3039	.002	-1.923	-.333
		4.0	-.7551*	.2676	.029	-1.455	-.055
	2.0	1.0	.7778*	.2606	.019	.096	1.459
		3.0	-.3500	.3255	.705	-1.201	.501
		4.0	.0227	.2919	1.000	-.741	.786

	3.0	1.0	1.1278*	.3039	.002	.333	1.923
		2.0	.3500	.3255	.705	-.501	1.201
		4.0	.3727	.3311	.675	-.493	1.239
	4.0	1.0	.7551*	.2676	.029	.055	1.455
		2.0	-.0227	.2919	1.000	-.786	.741
		3.0	-.3727	.3311	.675	-1.239	.493

*. The mean difference is significant at the 0.05 level.

The table displays the mean difference, standard error, significance level, and 95% confidence interval for each pair of groups being compared for three independent variables: Hosp_reg, Years_diag, and Bal_diet.

- For example, under the Hosp_reg variable, the mean difference between hospital visit with a good regularity (1.0) and those who never visit regularly (4.0) is significant at the 0.05 level, with a mean difference of -1.2475 and a confidence interval of [-2.155, -0.340]. This means that the mean score for the variable in question is significantly different between hospitals with regularity 1.0 and 4.0.
- Similarly, under the Years_diag variable, the mean difference between patients with a diagnosis of 1 year and those with a diagnosis of 4 years is significant at the 0.05 level, with a mean difference of 4.4141 and a confidence interval of [0.723, 8.105]. This means that the mean score for the variable in question is significantly different between patients with a diagnosis of 1 year and 4 years.
- Under the Bal_diet variable, the mean difference between patients with a balanced diet and those without a balanced diet is significant at the 0.05 level for all pairs of groups being compared, with mean differences ranging from -1.1278 to 0.7551 and confidence intervals that do not include 0. This means that the mean score for the variable in question is significantly different between patients with a balanced diet and those without a balanced diet, regardless of the level of registration or years since diagnosis.

Homogeneous Subsets

Medicine_reg

Tukey HSD[a,b]

Hosp_reg	N	Subset for alpha = 0.05	
		1	2
2.0	24	.583	
1.0	36	.889	
3.0	15	1.067	
4.0	22		2.136
Sig.		.596	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 22.031.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Years_diag

Tukey HSD[a,b]

Hosp_reg	N	Subset for alpha = 0.05	
		1	2
4.0	22	1.364	
3.0	15	3.800	3.800
2.0	24	4.500	4.500
1.0	36		5.778
Sig.		.197	.591

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 22.031.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Bal_diet

Tukey HSD[a,b]

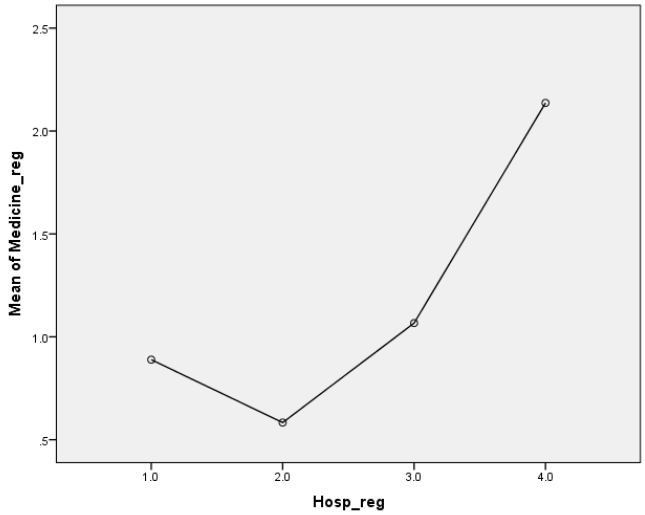
Hosp_reg	N	Subset for alpha = 0.05	
		1	2
1.0	36	1.472	
4.0	22	2.227	2.227
2.0	24	2.250	2.250
3.0	15		2.600
Sig.		.051	.596

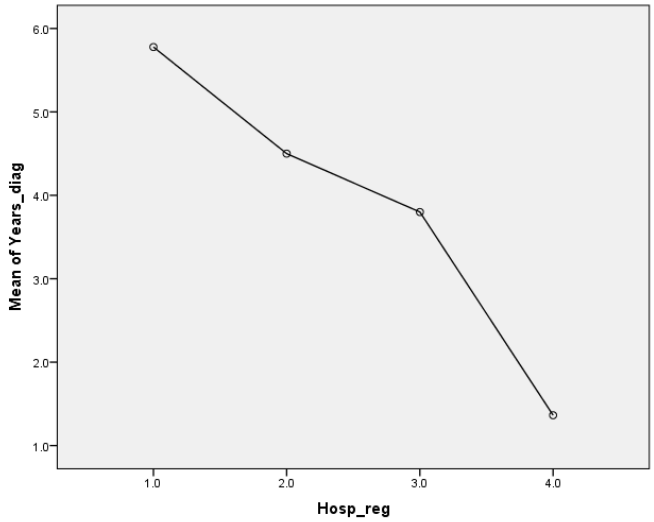
Means for groups in homogeneous subsets are displayed.

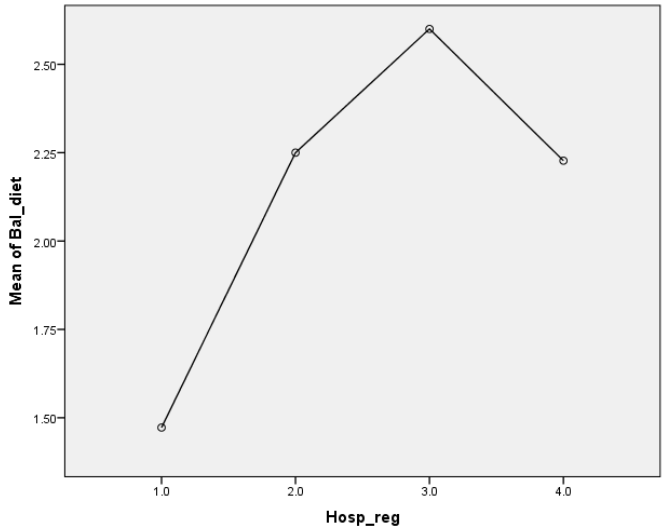
a. Uses Harmonic Mean Sample Size = 22.031.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Means Plots


Mean of Medicine_reg
2.5
2.0
1.5
1.0
.5
1.0
2.0
3.0
4.0
Hosp_reg


Mean of Years_diag
6.0
5.0
4.0
3.0
2.0
1.0
1.0
2.0
3.0
4.0
Hosp_reg


Mean of Bal_diet
2.50
2.25
2.00
1.75
1.50
1.0
2.0
3.0
4.0
Hosp_reg

One way

```
Weight Years_diag Bal_diet med_regularly maintain_BS_Level BY awareness_prog
```

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Weight	1.0	22	1.182	.5885	.1255	.921	1.443	.0	2.0
	2.0	8	1.625	.5175	.1830	1.192	2.058	1.0	2.0
	3.0	13	1.154	.6887	.1910	.738	1.570	.0	2.0
	4.0	54	1.278	.5636	.0767	1.124	1.432	.0	2.0
	Total	97	1.268	.5866	.0596	1.150	1.386	.0	2.0
Years_diag	1.0	22	2.818	3.7497	.7994	1.156	4.481	.0	14.0
	2.0	8	3.875	4.4861	1.5861	.125	7.625	.0	10.0
	3.0	13	7.923	8.5192	2.3628	2.775	13.071	.0	23.0
	4.0	54	3.833	4.8787	.6639	2.502	5.165	.0	20.0
	Total	97	4.155	5.3993	.5482	3.066	5.243	.0	23.0
Bal_diet	1.0	22	1.591	1.0538	.2247	1.124	2.058	1.0	4.0
	2.0	8	1.875	.8345	.2950	1.177	2.573	1.0	3.0
	3.0	13	1.846	1.0682	.2963	1.201	2.492	1.0	4.0
	4.0	54	2.241	1.0628	.1446	1.951	2.531	1.0	4.0
	Total	97	2.010	1.0655	.1082	1.796	2.225	1.0	4.0
med_regularly	1.0	22	1.500	.9636	.2054	1.073	1.927	1.0	4.0
	2.0	8	2.500	1.3093	.4629	1.405	3.595	1.0	4.0
	3.0	13	2.462	1.2659	.3511	1.697	3.227	1.0	4.0
	4.0	54	3.148	1.2348	.1680	2.811	3.485	1.0	4.0
	Total	97	2.629	1.3488	.1369	2.357	2.901	1.0	4.0
maintain_BS_Level	1.0	22	1.545	.9625	.2052	1.119	1.972	1.0	4.0
	2.0	8	2.000	1.0690	.3780	1.106	2.894	1.0	4.0
	3.0	13	2.077	1.0377	.2878	1.450	2.704	1.0	4.0
	4.0	54	2.593	1.1577	.1575	2.277	2.909	1.0	4.0
	Total	97	2.237	1.1617	.1179	2.003	2.471	1.0	4.0

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Weight	Between Groups	1.358	3	.453	1.329	.270
	Within Groups	31.673	93	.341		
	Total	33.031	96			
Years_diag	Between Groups	230.110	3	76.703	2.777	.046
	Within Groups	2568.571	93	27.619		
	Total	2798.680	96			
Bal_diet	Between Groups	7.234	3	2.411	2.204	.093
	Within Groups	101.756	93	1.094		
	Total	108.990	96			
med_regularly	Between Groups	43.094	3	14.365	10.155	.000
	Within Groups	131.546	93	1.414		
	Total	174.639	96			
maintain_BS_Level	Between Groups	18.132	3	6.044	5.045	.003
	Within Groups	111.415	93	1.198		
	Total	129.546	96			

Post Hoc Tests

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) awareness_prog	(J) awareness_prog	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Weight	1.0	2.0	-.4432	.2409	.262	-1.074	.187
		3.0	.0280	.2042	.999	-.506	.562
		4.0	-.0960	.1476	.915	-.482	.290
	2.0	1.0	.4432	.2409	.262	-.187	1.074
		3.0	.4712	.2622	.281	-.215	1.157
		4.0	.3472	.2211	.400	-.231	.926
	3.0	1.0	-.0280	.2042	.999	-.562	.506
		2.0	-.4712	.2622	.281	-1.157	.215
		4.0	-.1239	.1803	.902	-.596	.348

	4.0	1.0	.0960	.1476	.915	-.290	.482
		2.0	-.3472	.2211	.400	-.926	.231
		3.0	.1239	.1803	.902	-.348	.596
Years_diag	1.0	2.0	-1.0568	2.1697	.962	-6.733	4.619
		3.0	-5.1049*	1.8385	.033	-9.914	-.295
		4.0	-1.0152	1.3292	.871	-4.493	2.462
	2.0	1.0	1.0568	2.1697	.962	-4.619	6.733
		3.0	-4.0481	2.3616	.322	-10.226	2.130
		4.0	.0417	1.9909	1.000	-5.167	5.250
	3.0	1.0	5.1049*	1.8385	.033	.295	9.914
		2.0	4.0481	2.3616	.322	-2.130	10.226
		4.0	4.0897	1.6236	.064	-.158	8.337
	4.0	1.0	1.0152	1.3292	.871	-2.462	4.493
		2.0	-.0417	1.9909	1.000	-5.250	5.167
		3.0	-4.0897	1.6236	.064	-8.337	.158
Bal_diet	1.0	2.0	-.2841	.4319	.913	-1.414	.846
		3.0	-.2552	.3659	.898	-1.213	.702
		4.0	-.6498	.2646	.074	-1.342	.042
	2.0	1.0	.2841	.4319	.913	-.846	1.414
		3.0	.0288	.4700	1.000	-1.201	1.259
		4.0	-.3657	.3963	.793	-1.402	.671
	3.0	1.0	.2552	.3659	.898	-.702	1.213
		2.0	-.0288	.4700	1.000	-1.259	1.201
		4.0	-.3946	.3232	.615	-1.240	.451
	4.0	1.0	.6498	.2646	.074	-.042	1.342
		2.0	.3657	.3963	.793	-.671	1.402
		3.0	.3946	.3232	.615	-.451	1.240
med_regularly	1.0	2.0	-1.0000	.4910	.182	-2.285	.285
		3.0	-.9615	.4161	.103	-2.050	.127
		4.0	-1.6481*	.3008	.000	-2.435	-.861
	2.0	1.0	1.0000	.4910	.182	-.285	2.285
		3.0	.0385	.5344	1.000	-1.360	1.437
		4.0	-.6481	.4506	.479	-1.827	.531
	3.0	1.0	.9615	.4161	.103	-.127	2.050
		2.0	-.0385	.5344	1.000	-1.437	1.360
		4.0	-.6866	.3674	.249	-1.648	.275
	4.0	1.0	1.6481*	.3008	.000	.861	2.435

		2.0	.6481	.4506	.479	-.531	1.827
		3.0	.6866	.3674	.249	-.275	1.648
maintain_BS_Level	1.0	2.0	-.4545	.4519	.746	-1.637	.728
		3.0	-.5315	.3829	.510	-1.533	.470
		4.0	-1.0471*	.2768	.002	-1.771	-.323
	2.0	1.0	.4545	.4519	.746	-.728	1.637
		3.0	-.0769	.4918	.999	-1.364	1.210
		4.0	-.5926	.4147	.485	-1.677	.492
	3.0	1.0	.5315	.3829	.510	-.470	1.533
		2.0	.0769	.4918	.999	-1.210	1.364
		4.0	-.5157	.3381	.427	-1.400	.369
	4.0	1.0	1.0471*	.2768	.002	.323	1.771
		2.0	.5926	.4147	.485	-.492	1.677
		3.0	.5157	.3381	.427	-.369	1.400

*. The mean difference is significant at the 0.05 level.

Homogeneous Subsets

Weight

Tukey HSD[a,b]

awareness_prog	N	Subset for alpha = 0.05
		1
3.0	13	1.154
1.0	22	1.182
4.0	54	1.278
2.0	8	1.625
Sig.		.127

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 15.043.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Years_diag

Tukey HSD[a,b]

awareness_prog	N	Subset for alpha = 0.05	
		1	2
1.0	22	2.818	
4.0	54	3.833	3.833
2.0	8	3.875	3.875
3.0	13		7.923
Sig.		.946	.150

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 15.043.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Bal_diet

Tukey HSD[a,b]

awareness_prog	N	Subset for alpha = 0.05
		1
1.0	22	1.591
3.0	13	1.846
2.0	8	1.875
4.0	54	2.241
Sig.		.328

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 15.043.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

med_regularly

Tukey HSD[a,b]

awareness_prog	N	Subset for alpha = 0.05	
		1	2
1.0	22	1.500	
3.0	13	2.462	2.462
2.0	8	2.500	2.500
4.0	54		3.148
Sig.		.104	.393

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 15.043.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

maintain_BS_Level

Tukey HSD[a,b]

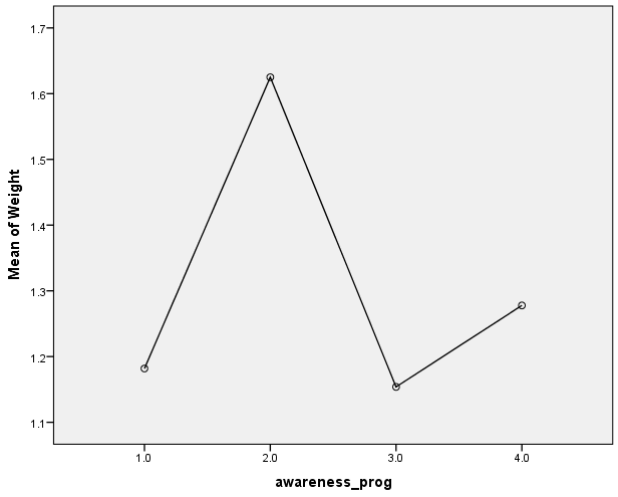
awareness_prog	N	Subset for alpha = 0.05	
		1	2
1.0	22	1.545	
2.0	8	2.000	2.000
3.0	13	2.077	2.077
4.0	54		2.593
Sig.		.545	.451

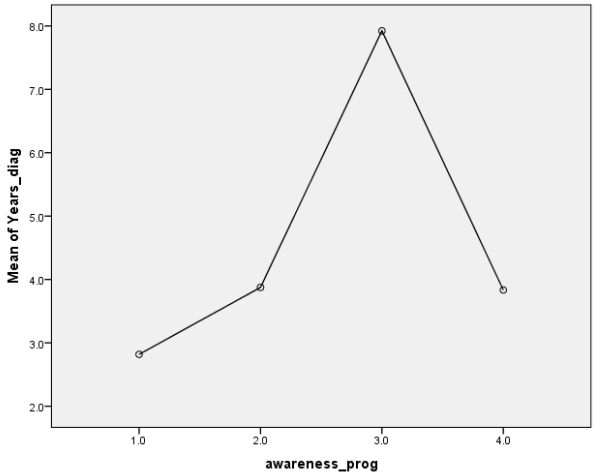
Means for groups in homogeneous subsets are displayed.

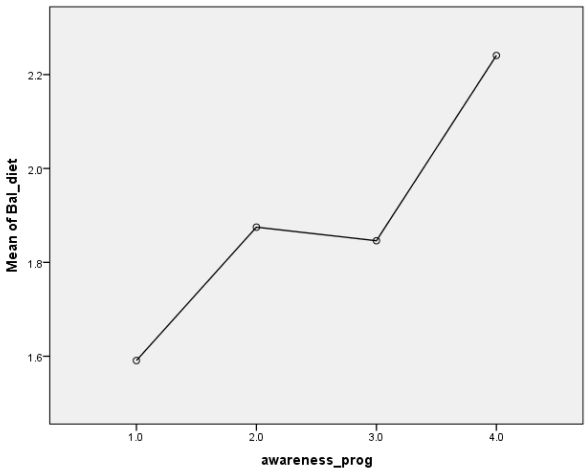
a. Uses Harmonic Mean Sample Size = 15.043.

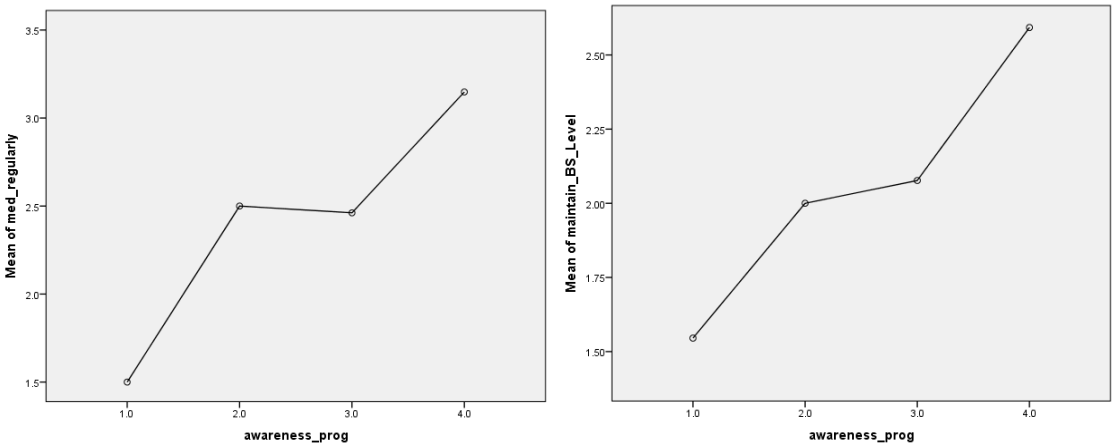
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Means Plots









The output shows the descriptive statistics for each variable in the dataset, as well as the results of conducting a one-way ANOVA with post-hoc Tukey's HSD test.

The descriptive statistics provide information on the central tendency and variability of each variable. For example, we can see that the average weight is 1.268, with a standard deviation of 0.5866, and that the range of weight is from 0 to 2.

The ANOVA tests whether there is a statistically significant difference in means between the groups for each variable. The null hypothesis is that there is no difference between the means of the groups, while the alternative hypothesis is that at least one group's mean is different from the others. The ANOVA table provides information on the F-statistic, degrees of freedom, and p-value for each variable.

The post-hoc Tukey's HSD test is conducted when the ANOVA test shows a statistically significant difference between groups. This test compares the means of all possible pairs of groups to determine which ones are significantly different from each other. The output shows the p-values for each pairwise comparison, as well as the adjusted p-value (alpha) for the entire test.

General Linear Model

```
GLM Hereditary BP BY Type WITH Weight
  /WSFACTOR=factor1 2 Polynomial
  /METHOD=SSTYPE (3)
  /CRITERIA=ALPHA (.05)
  /WSDESIGN=factor1
  /DESIGN=Weight Type.
```

Within-Subjects Factors

Measure: MEASURE_1

factor1	Dependent Variable
1	Hereditary
2	BP

Between-Subjects Factors

		N
Type	1.0	44
	2.0	53

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
factor1	Pillai's Trace	.046	4.506[b]	1.000	94.000	.036
	Wilks' Lambda	.954	4.506[b]	1.000	94.000	.036
	Hotelling's Trace	.048	4.506[b]	1.000	94.000	.036
	Roy's Largest Root	.048	4.506[b]	1.000	94.000	.036
factor1 * Weight	Pillai's Trace	.048	4.734[b]	1.000	94.000	.032
	Wilks' Lambda	.952	4.734[b]	1.000	94.000	.032
	Hotelling's Trace	.050	4.734[b]	1.000	94.000	.032
	Roy's Largest Root	.050	4.734[b]	1.000	94.000	.032
factor1 * Type	Pillai's Trace	.003	.268[b]	1.000	94.000	.606
	Wilks' Lambda	.997	.268[b]	1.000	94.000	.606
	Hotelling's Trace	.003	.268[b]	1.000	94.000	.606
	Roy's Largest Root	.003	.268[b]	1.000	94.000	.606

a. Design: Intercept + Weight + Type

Within Subjects Design: factor1

b. Exact statistic

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
factor1	1.000	.000	0	.	1.000	1.000	1.000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + Weight + Type

Within Subjects Design: factor1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	2.072	1	2.072	4.506	.036
	Greenhouse-Geisser	2.072	1.000	2.072	4.506	.036
	Huynh-Feldt	2.072	1.000	2.072	4.506	.036
	Lower-bound	2.072	1.000	2.072	4.506	.036
factor1 * Weight	Sphericity Assumed	2.177	1	2.177	4.734	.032
	Greenhouse-Geisser	2.177	1.000	2.177	4.734	.032
	Huynh-Feldt	2.177	1.000	2.177	4.734	.032
	Lower-bound	2.177	1.000	2.177	4.734	.032
factor1 * Type	Sphericity Assumed	.123	1	.123	.268	.606
	Greenhouse-Geisser	.123	1.000	.123	.268	.606
	Huynh-Feldt	.123	1.000	.123	.268	.606
	Lower-bound	.123	1.000	.123	.268	.606
Error(factor1)	Sphericity Assumed	43.221	94	.460		
	Greenhouse-Geisser	43.221	94.000	.460		
	Huynh-Feldt	43.221	94.000	.460		
	Lower-bound	43.221	94.000	.460		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	factor1	Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Linear	2.072	1	2.072	4.506	.036
factor1 * Weight	Linear	2.177	1	2.177	4.734	.032
factor1 * Type	Linear	.123	1	.123	.268	.606
Error(factor1)	Linear	43.221	94	.460		

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	10.271	1	10.271	20.434	.000
Weight	.256	1	.256	.510	.477
Type	.024	1	.024	.047	.829
Error	47.246	94	.503		

This output is from a General Linear Model (GLM) analysis. The analysis includes one within-subjects factor (factor1) with two levels (Hereditary and BP) and two between-subjects factors (Weight and Type). The dependent variable is not specified in the output.

The between-subjects factor "Type" has two levels: 1.0 and 2.0, with sample sizes of 44 and 53, respectively. The between-subjects factor "Weight" has a continuous scale.

The multivariate tests table shows the results of the multivariate analysis of variance (MANOVA) for the within-subjects and between-subjects factors. The multivariate tests show that the effect of factor1 (main effect of within-subjects factor) and factor1Weight (interaction effect of within-subjects and between-subjects factors) are significant, with Pillai's Trace values of .046 and .048, respectively, and p-values of .036 and .032, respectively. However, the effect of factor1Type is not significant (p = .606).
The Mauchly's test of sphericity output suggests that the assumption of sphericity is met (p=1.00) for the within-subjects factor.

The tests of within-subjects effects table show the results of the univariate ANOVAs for the within-subjects and between-subjects factors. The output shows that there is a significant effect of factor1 and the interaction between factor1 and Weight. The effect of the interaction between factor1 and Type is not significant. The Greenhouse-Geisser correction was used to adjust the degrees of freedom for the averaged tests of significance.

CONCLUSION

Based on the analysis of our dataset, we can conclude that the majority of the patients (59.8%) in our sample were female, with 40.2% being male. In terms of age distribution, most patients were above 55 years old (43.4%), followed by those aged 41-55 (15.2%), 26-40 (24.2%), and 10-25 (2.0%). The median years since diagnosis was found to be 5.0 years, with the majority of patients being diagnosed within the past 10 years (88.7%). The remaining patients were diagnosed between 10-23 years ago (3.1%) and 23 or more years ago (1.0%). In terms of type of diagnosis, the majority of patients had Type 2 diabetes (54.6%), with the remaining patients having Type 1 diabetes (45.4%).

Overall, these findings provide valuable insights into the demographics and characteristics of patients with diabetes in our sample. The information can be used to guide future research, improve healthcare policies and practices, and ultimately lead to better outcomes for individuals with diabetes.

REFERENCES:

1. Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H. Predicting Diabetes Mellitus With Machine Learning Techniques. Front Genet. 2018 Nov 6;9:515. doi: 10.3389/fgene.2018.00515. PMID: 30459809; PMCID: PMC6232260.
2. M. K. Hasan, M. A. Alam, D. Das, E. Hossain and M. Hasan, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers," in IEEE Access, vol. 8, pp. 76516-76531, 2020, doi:10.1109/ACCESS.2020.2989857. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9076634
3. Varun Jaiswal, Anjli Negi, Tarun Pal,A review on current advances in machine learning based diabetes prediction,Primary Care Diabetes, Volume 15, Issue 3,2021,Pages 435-443,ISSN 1751-9918, https://doi.org/10.1016/j.pcd.2021.02.005
4. N.M. Saravana kumar, T. Eswari, P. Sampath, S. Lavanya,Predictive Methodology for Diabetic Data Analysis in Big Data,Procedia Computer Science,Volume 50,2015,Pages 203-208,ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.04.069
5. Talha Mahboob Alam, Muhammad Atif Iqbal, Yasir Ali, Abdul Wahab, Safdar Ijaz, Talha Imtiaz Baig, Ayaz Hussain, Muhammad Awais Malik, Muhammad Mehdi Raza, Salman Ibrar, Zunish Abbas,A model for early prediction of diabetes,Informatics in Medicine Unlocked,Volume 16,2019,100204,ISSN 2352-9148, https://doi.org/10.1016/j.imu.2019.100204
6. Aishwarya Mujumdar, V Vaidehi,Diabetes Prediction using Machine Learning Algorithms,Procedia Computer Science,Volume 165,2019,Pages 292-299,ISSN 1877-0509,https://doi.org/10.1016/j.procs.2020.01.047
7. M. Paliwal and P. Saraswat, "Research on Diabetes Prediction Method Based on Machine Learning," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 415-419, doi: 10.1109/ICTACS56270.2022.9988050.https://iopscience.iop.org/article/10.1088/1742-6596/1684/1/012062/pdf
8. Rani, KM. (2020). Diabetes Prediction Using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 294-305. 10.32628/CSEIT206463. https://www.researchgate.net/publication/347091823_Diabetes_Prediction_Using_Machine_Learning
9. Ayon, Safial & Islam, Md. (2019). Diabetes Prediction: A Deep Learning Approach. International Journal of Information Engineering and Electronic Business. 11. 21-27. 10.5815/ijieeb.2019.02.03. https://www.researchgate.net/publication/332298424_Diabetes_Prediction_A_Deep_Learning_Approach#:~:text=The%20results%20on%20PID%20dataset,five%2Dfold%20cross%2Dvalidation
10. Naz H, Ahuja S. Deep learning approach for diabetes prediction using PIMA Indian dataset. J Diabetes Metab Disord. 2020 Apr 14;19(1):391-403. doi: 10.1007/s40200-020-00520-5. PMID: 32550190; PMCID: PMC7270283.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7270283/
11. https://www.liebertpub.com/doi/pdf/10.1089/pop.2018.0129
12. https://www.binasss.sa.cr/medint/ART07.pdf
13. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine Learning and Data Mining Methods in Diabetes Research. Comput Struct Biotechnol J. 2017 Jan 8;15:104-116. doi: 10.1016/j.csbj.2016.12.005. PMID: 28138367; PMCID: PMC5257026. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5257026/
14. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res. 2018 May 30;20(5):e10775. doi: 10.2196/10775. PMID: 29848472; PMCID: PMC6000484. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6000484/

15. Dinh, A., Miertschin, S., Young, A. et al. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak 19, 211 (2019). https://doi.org/10.1186/s12911-019-0918-5
16. Maniruzzaman, M., Rahman, M.J., Ahammed, B. et al. Classification and prediction of diabetes disease using machine learning paradigm. Health Inf Sci Syst 8, 7 (2020). https://doi.org/10.1007/s13755-019-0095-z
17. Zhou, H., Myrzashova, R. & Zheng, R. Diabetes prediction model based on an enhanced deep neural network. J Wireless Com Network 2020, 148 (2020). https://doi.org/10.1186/s13638-020-01765-7
18. Al Yousef MZ, Yasky AF, Al Shammari R, Ferwana MS. Early prediction of diabetes by applying data mining techniques: A retrospective cohort study. Medicine (Baltimore). 2022 Jul 22;101(29):e29588. doi: 10.1097/MD.0000000000029588. PMID: 35866773; PMCID: PMC9302319. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9302319/#:~:text=Their%20model%20achieved%2094%25%20accuracy,%2C%20they%20excluded%20pre%2Ddiabetics
19. Rashi Rastogi, Mamta Bansal,Diabetes prediction model using data mining techniques,Measurement: Sensors,Volume 25,2023,100605,ISSN 2665-9174,https://doi.org/10.1016/j.measen.2022.100605
20. Jayasri N.P., R. Aruna,Big data analytics in health care by data mining and classification techniques,ICT Express,Volume 8, Issue 2,2022,Pages 250-257,ISSN 2405-9595,https://doi.org/10.1016/j.icte.2021.07.001
21. Diabetes Data Prediction in healthcare Using Hadoop over Big Data. European Journal of Molecular & Clinical Medicine, 7(4), 1423-1432. https://ejmcm.com/article_1840.html
22. Metsker, O., Magoev, K., Yakovlev, A. et al. Identification of risk factors for patients with diabetes: diabetic polyneuropathy case study. BMC Med Inform Decis Mak 20, 201 (2020). https://doi.org/10.1186/s12911-020-01215-w
23. G. Woldemichael and S. Menaria, "Prediction of Diabetes Using Data Mining Techniques," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2018, pp. 414-418, doi: 10.1109/ICOEI.2018.8553959. https://ieeexplore.ieee.org/document/8553959
24. https://www.scirp.org/pdf/jcc_2022110114390330.pdf
25. F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab and S. A. C. Bukhari, "Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review," in IEEE Access, vol. 9, pp. 43711-43735, 2021, doi: 10.1109/ACCESS.2021.3059343. https://ieeexplore.ieee.org/document/9354154
26. Larabi-Marie-Sainte, S.; Aburahmah, L.; Almohaini, R.; Saba, T. Current Techniques for Diabetes Prediction: Review and Case Study. Appl. Sci. 2019, 9, 4604. https://doi.org/10.3390/app9214604