Time Series Monitoring Data

Group 13

SIMPAO, CHARIZE R.
TANYAG, LORD EXZEL JHONNE L.

INTRODUCTION

This report is a study of time series monitoring data contained in Time series Mointoring data. The data traces mean number of steps, stress, and BMI of three separate patients in 42 months. The main goal is to investigate how these variables are related to each other across the time and develop a model that could predict the stress level of Patient 1 depending on his/her BMI and average steps. The preprocessing process involved removing the column marked as a month and deleting any given row with missing values because it could interfere with consistent data analysis.

METHODOLOGY

- Data Cleaning
- Summary Statistics
- Data Structure Inspection
- Correlation Matrix
- Histograms
- Boxplots

Multiple Linear Regression (MLR): linear regression model was built to predict Patient_1_Stress_Level using Patient_1_BMI and Patient_1_avg_steps as predictors.
- The summary() function provided details on coefficients, R-squared values, and statistical significance.
- A scatter plot comparing the predicted vs. actual stress levels for Patient 1 was generated to assess the model's accuracy.

Categorical Analysis (BMI and Steps Groups):

This report is a study of time series monitoring data contained in Time series Mointoring data. The data traces mean number of steps, stress, and BMI of three separate patients in 42 months. The main goal is to investigate how these variables are related to each other across the time and develop a model that could predict the stress level of Patient 1 depending on his/her BMI and average steps. The preprocessing process involved removing the column marked as a month and deleting any given row with missing values because it could interfere with consistent data analysis.

Results and Figures

Data Summary

The dataset contains information on the lifetime, or life cycle, of a set of programming classes that were used by a research group in testing. Inside the dataset, there are three files: the Summar, the LA, and the Lifetime. In the Summar, there is named a set of classes that are tested by a research group. In the LA that is located in the dataset, the lifetime of these classes is described. In the Lifetime that is located inside the dataset, there are the descriptions of the lifetime of the classes that are tested by a research group.

The resulting dataset is composed of 42 observations and 10 variables with the numerical representations of months, average steps, stress levels, and BMI of three various patients. These variables have been summed up briefly in table 1. As an example, Patient1avgsteps will be between 302 and 8542 with an average of around 3691 steps. Each patient has a stress level between 1 and 10. The BMI values of the Patient 1 are 23-26, Patient 2 24-24.9 and Patient 3 23.3-26.3.

Table 1. Summary Statistics of Numeric Variables in the Cleaned Time Series Monitoring Dataset.

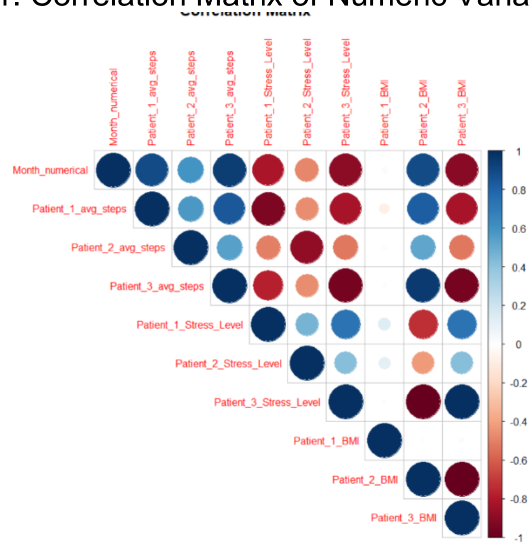
Statistic	Month_numerical	Patient_1_avg_steps	Patient_2_avg_steps	Patient_3_avg_steps
Min.	1	302	578	273
1st Qu.	11.25	1368	1614	1689
Median	21.5	3574	2562	3822
Mean	21.5	3691	2839	4259
3rd Qu.	31.75	5283	4265	6560
Max.	42	8542	5956	9833

Patient_1_Stress_Level	Patient_2_Stress_Level	Patient_3_Stress_Level
2	3	1
4.25	6	3
7	7	6
6.429	6.905	5.548
8	8	8
10	10	10
Patient_1_BMI	Patient_2_BMI	Patient_3_BMI
23	24	23.3
24	24.2	24
25	24.4	25
24.81	24.45	24.85
26	24.7	25.7
26	24.9	26.3

VISUALIZATIONS

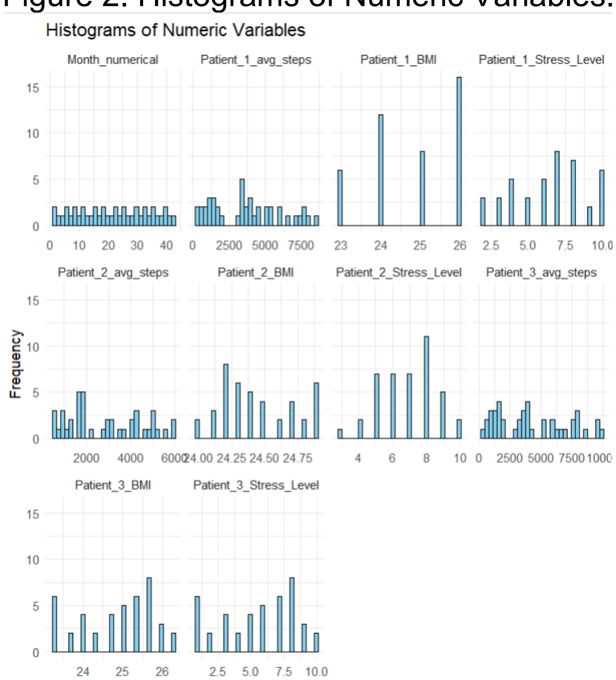
This correlation matrix will show how the variables are interrelated with each other in a linear way. Positive associations are high among the average steps, stress levels, and the BMI of each patient. As an example Patient_1_Stress_Level is strongly positive correlated with Patient_1_BMI. Also, there can be different levels of correlation of metrics between different patients. This circle plot displays the correlation table of all the numeric variables present in the data. The magnitude and the direction of the correlation is given by the size of each circle and the color: darker blue means a higher positive correlation, darker red a higher negative correlation and lighter colours mean weaker correlations. (Figure 1).

Figure 1. Correlation Matrix of Numeric Variables.



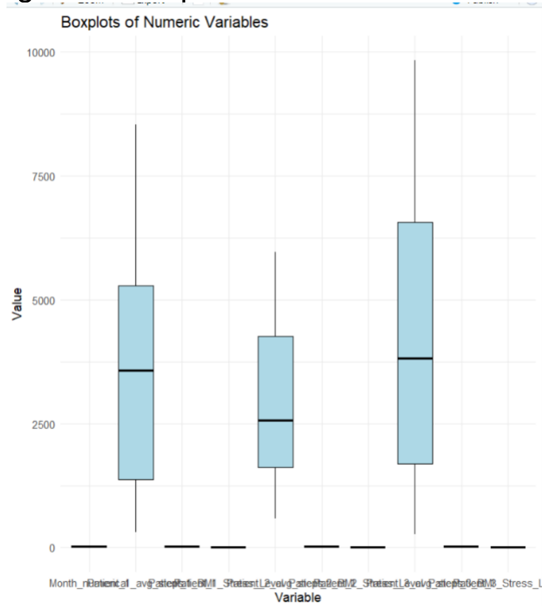
The histograms represent the distributions of each numeric variables. Numerous variables, for example, average steps per patient, exhibit various peaks or more or less constant distribution, which indicates variability over time in months. The BMI values of the individual patients seem deliberately confined to small variations, which shows that they did not change much over time. Each patient has several amounts of maximums in their stress range instead of one normal distribution. It is a multi panel figure which shows the frequency distribution of different numeric variables. The histograms are based on 30 bins (default) that depict the frequency of entries in each range, and the bars are painted in light blue and have black outlines. Each subplot utilizes the free scaling of the x-axis to cover the range of the given individual variable most effectively. (Figure 2).

Figure 2. Histograms of Numeric Variables.



There are broad ranges as well as outliers in average steps among all the patients suggesting poor consistency of physical activity. Interquartile ranges of BMI values, especially in Patient 2, are very small, which points to high consistency. The distribution of stress levels is normally medium. (Figure 3).

Figure 3. Boxplots of Numeric Variables.

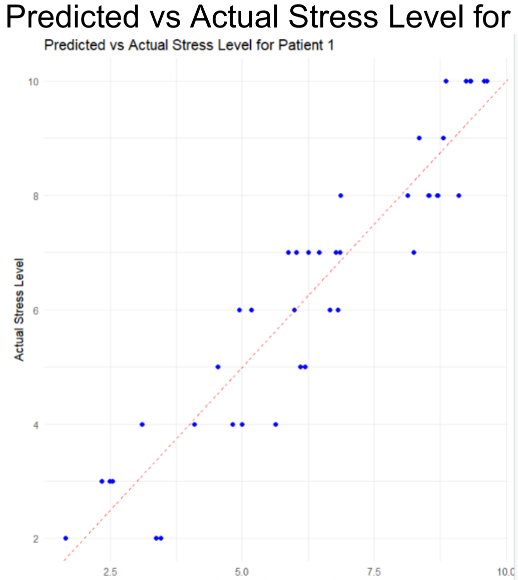


Linear Regression Model for Patient 1's Stress Level

- The R-squared statistic was basically 0.8765 which means that about 87.65 percent of the variability of the stress level of patient 1 can be attributed to these two predictors.
- Displays the estimated coefficients for patient_1_BMI and Patient_1_avg_steps in the linear regression model predicting Patient_1_Stress_Level. Standard errors, t-values, and p-values are also provided. Significance codes indicate the statistical significance of each coefficient, with * denoting p < 0.001 and * denoting p < 0.05.

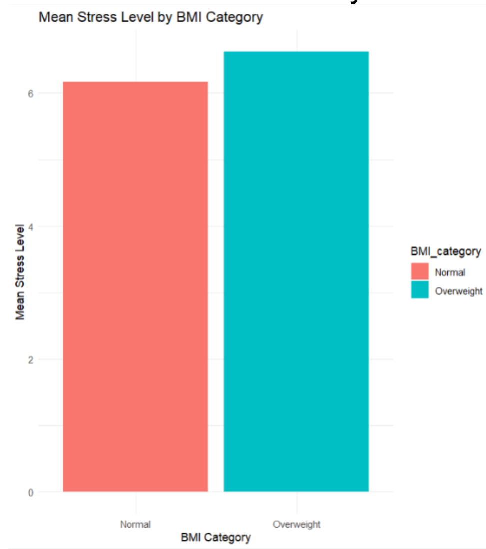
This scatter plot demonstrates the estimates of the stress levels and the actual stress levels of patient 1. The points are packed around the dashed red curve independent of normal (suggesting that the model is a good matching exercise in gauging the stress level of Patient 1). This scatter-plot is a comparison of the predicted and real arrival Stress Level based on linear regression model of patient 1. Markers are in blue and each data point is represented by a marker, with a dashed line that shows perfect prediction, i.e. slope = 1 and intercept = 0 (y=x) in red. The high correlation of datapoints towards this line proves the high predictive quality of the model. (Figure 4).

Figure 4. Predicted vs Actual Stress Level for Patient 1.



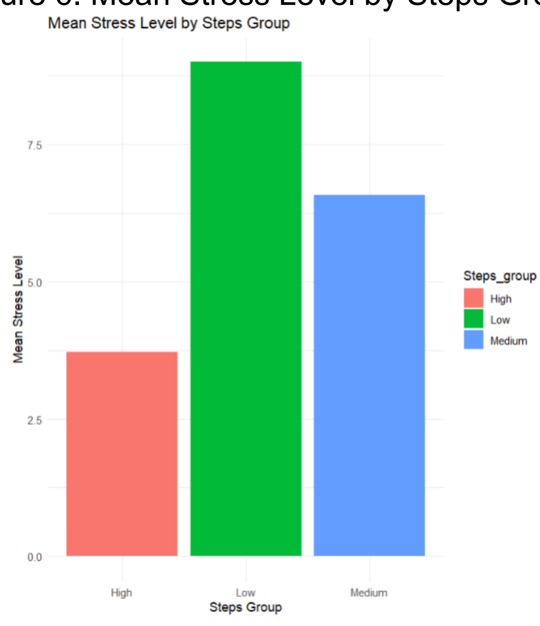
This is a bar plot that is used to compare the means of stress level (of Patient 1) in two categories of BMI: Normal and Overweight. The overweight category displays a slightly increased mean stress level than in the normal category in the BMI. (Figure 5)

Figure 5. Mean Stress Level by BMI Category.



This bar plot shows the mean level of stress of Patient 1 at three levels of average steps: This confirms the inverse relationship presented in the regression model. (Figure 6).

Figure 6. Mean Stress Level by Steps Group.



CONCLUSION

To sum up, this analysis proves the close connection between physical activity (average steps) and stress levels. These findings might lead to the idea of promoting the active lifestyle as a stress management tool. Although the direct linear association of BMI with stress did not turn out to be significant in Patient 1, there is a glimpse of it when it comes to a categorical picture. A more advanced study would be the possibility of examining non-linear relationships, all analysis with time-lan effects, and individual deviation of patients in more depth.