Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 27, Issue 2 (2025) ISSN$_{Print}$: 1816-9171
DOI: https:// doi.org/ 10.33916/ qjae.2025 .02112117

Factorial Experimental Analysis of Factors Affecting Hypertension

Mohammed H. AL- Sharoot Mohammed O. AL-Katib
mohammed.alsharoot@qu.edu.iq ma7767@ntu.edu.iq
University of Al-Qadisiyah

Article history:
Received: 22/12/2024
Accepted: 24/12/2024
Available online: 15 /6 /2025

Corresponding Author : Mohammed O. AL-Katib

Abstract: Hypertension (Blood pressure) is considered one of the world's most important indicators of cardiovascular disease. Therefore, attention is focused on studying the most important factors that cause high blood pressure in order to spread health awareness to individuals to avoid this disease. Factorial design is a widely used statistical tool in many scientific disciplines that aims to determine the effects of factors with levels and their interaction on the experimental response units to those factors. In this paper, we mainly focus on using adaptive lasso along with factorial analysis to study the selecting of the most influential factors and interactions that affect hypertension. Hence, the results show that the adaptive Lasso refines the model by selecting only the most impactful variables, which enhances model interpretation ability and accuracy.

Keyword: Factorial Experimental , Regression , Adaptive Lasso , Hypertension , Kolmogorov-Smirnov , ShapiroWilk.

INTRODUCTION: Hypertension, or high blood pressure, is one of the most prevalent indicators of cardiovascular disease worldwide, posing serious health risks and leading to various complications if left unmanaged. Given its global impact, there is a critical need to understand the factors that contribute to hypertension to improve prevention and health awareness [1]. Identifying these key factors allows for more targeted interventions, potentially reducing the incidence of hypertension-related complications. Factorial design is a robust statistical tool widely employed in numerous scientific fields to study the effects of multiple factors and their interactions on outcomes [2]. This approach is especially valuable in understanding complex conditions like hypertension, where multiple lifestyle, genetic, and environmental factors may contribute simultaneously [3]. By structuring experiments with factorial design, researchers can assess the individual influence of each factor and observe how these factors interact, leading to insights into combined effects that may amplify or reduce hypertension risk.

In studying high-dimensional data, challenges arise when attempting to select the most influential factors while avoiding overfitting and multicollinearity issues [4]. To address this, recent advancements in variable selection methods, such as the adaptive lasso, have become invaluable. Adaptive lasso enables efficient selection of significant factors by penalising less impactful variables, thus refining the model to focus on key drivers of hypertension. This method is particularly useful in factorial experiments, where multiple factors and interactions can result in a complex, high-dimensional data structure. This paper integrates adaptive lasso with factorial analysis to examine and select the most influential factors and interactions impacting hypertension. By employing this combined approach, we aim to provide a clearer understanding of the multifaceted causes of high blood pressure, offering a foundation for more effective interventions and health awareness campaigns. A full factorial experimental design effectively examines the effects of multiple factors and their interactions on a response variable. In this design, all possible combinations of factor levels are included, allowing for a comprehensive analysis of both main effects and interactions. When each factor in the design has only two levels, the experimental setup is often referred to as a 2^kFactorial design, where k represents the number of factors [3]. This structure is particularly useful in identifying which factors have a significant impact on the outcome, as well as how combinations of factors interact to influence the response [5].

In the context of studying hypertension, we utilize a full factorial design with two levels for each factor, representing conditions such as "high" and "low" or "presence" and "absence." For example, factors such as salt intake, physical activity, and stress can each be assigned two levels (e.g., high vs. low salt intake, active vs. sedentary lifestyle) [6] .

This setup allows us to explore how each individual factor influences blood pressure and identify possible interaction effects between factors. In a 2^kfactorial design, each factor has exactly two levels, and all combinations of these levels across the k factors are tested. This results in 2^kexperimental conditions or runs. For instance [3], with three factors, a 2^3 design would yield eight experimental conditions, covering every possible combination of the levels across these factors [2]. The factorial design setup for this study includes six factors, each with two levels. The total number of experimental conditions is therefore $2^k = 64$. Each of these 64 conditions represents a unique combination of the six factors' levels, allowing for a detailed examination of both main effects and interactions up to six-way interactions [1].
The full factorial design offers several advantages in the study of hypertension [7]: Comprehensive Analysis: By including all possible combinations of factor levels, the design enables a complete assessment of each factor's effect on hypertension and allows for the exploration of interactions between factors. Interaction Effects: Interaction effects, especially between lifestyle and genetic factors, can provide critical insights into how combinations of behaviours or predispositions contribute to blood pressure changes [6]. These interactions would be difficult to capture without a factorial approach. Efficient Use of Data: Although factorial designs may require a large number of runs, they are highly efficient for studying multiple factors simultaneously, making them well-suited for complex health studies with multifactorial influences.
For practical analysis, each factor level is coded as either (+1) or (1), representing high and low levels, respectively. This coding system simplifies the analysis by transforming categorical factors into numerical values, allowing for regression and other statistical analyses [8]. For example, High salt intake: (+1), Low salt intake: (1) , this binary coding helps create the design matrix (X), which represents the full factorial structure of the experiment. Each row in (X) corresponds to one experimental run, with columns representing the coded levels of each factor. By employing this full factorial design with two levels, we can systematically examine the effects of each factor and their interactions on blood pressure, providing a foundation for the adaptive lasso analysis in subsequent sections. In the context of this study, EER and IER provide complementary assessments of the model's quality [7]. EER evaluates the model’s overall predictive power, while IER focuses on its ability to capture the intricate interplay between factors. Given the multifactorial nature of hypertension, where lifestyle, genetic, and environmental factors interact to influence blood pressure, it is crucial to optimize both EER and IER to ensure the model is both accurate and interpretable.
These criteria guide the selection and validation of factors and interaction terms in the model, enabling the identification of the most influential variables on hypertension. By minimizing EER and IER, we can ensure that the adaptive lasso model reliably identifies the key drivers and interactions affecting blood pressure, providing a foundation for more targeted hypertension interventions and recommendations. In factorial experiments, particularly those with multiple factors and interactions, variable selection is essential for creating a manageable and interpretable model. High-dimensional data often includes numerous factors and interactions, some of which may not significantly impact the outcome [11]. Effective variable selection helps eliminate less influential variables, reducing noise and focusing the model on the most impactful factors. In this study, we employ the adaptive Lasso method to address variable selection within the factorial design framework for analyzing factors affecting hypertension. In this study, we integrate adaptive Lasso with factorial design to identify the most significant factors and interactions affecting hypertension. Each factor in our factorial design has two levels (e.g., high vs. low salt intake, active vs. sedentary lifestyle), and adaptive Lasso helps focus the model on the main effects and interactions that have the strongest influence on blood pressure. The adaptive nature of this method makes it particularly suitable for factorial designs with many interactions, as it applies varying levels of penalty based on the initial influence of each factor and interaction term. This selective penalization effectively reduces the model’s dimensionality by retaining only the most impactful variables, thus enhancing interpretability and focusing the analysis on key factors. Using adaptive Lasso within the factorial design framework allows us to build a refined model that highlights the primary drivers of high blood pressure [16]. By focusing on significant main effects and interactions, adaptive Lasso enables us to achieve a clearer understanding of the factors influencing hypertension, supporting the development of more effective health interventions.

1. Regression Model for Factorial Design

In factorial experiments, it is essential to analyze the relationship between factors and the response variable, in this case, blood pressure levels. The regression model provides a framework for evaluating both the main effects of individual factors and their interactions [2]. By incorporating the factorial design into a regression framework, we can quantify the contribution of each factor and interaction term to hypertension, identifying significant influences with precision.

1.1. Structure of the Regression Model

Given a factorial design with (k) factors, each with two levels, the general form of the regression model for a full factorial 2^k design is[3]:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{i<j} \beta_{ij} X_i X_j + \sum_{i<j<l} \beta_{ijl} X_i X_j X_l + \cdots + \varepsilon \qquad (1)$$

where:

(Y) represents the response variable (blood pressure level), (β_0) is the intercept, (β_i) represents the main effect coefficients for each factor, (β_{ij}), (β_{ijl}), etc., are the coefficients for two-factor, three-factor, and higher-order interactions, respectively, (X_i) is the coded level ((+1) or (1)) of factor (i), (ε) is the error term, assumed to be normally distributed with mean zero and constant variance. This model accounts for the influence of each factor on blood pressure, as well as the combined effects when multiple factors are present together. By including interaction terms, we can capture the synergistic effects that may exist between factors, which are often crucial in medical studies like hypertension analysis.

The regression coefficients (β) are estimated using least squares methods, where the objective is to minimize the sum of squared residuals [8]. In the context of factorial design, the coded levels ((+1) and (1)) of the factors simplify the estimation process, making it straightforward to interpret each coefficient as the average effect of increasing the factor from its low to high level. Positive values of (β) indicate that increasing the level of the factor (from low to high) is associated with an increase in blood pressure, while negative values suggest a decrease. Interaction coefficients indicate how pairs or groups of factors jointly influence blood pressure. For example, a significant positive interaction between salt intake and stress may imply that high levels of both factors result in a compounded increase in blood pressure.

1.2. Evaluation Criteria EER and IER

In analysing regression models, particularly in the context of factorial design with high-dimensional data, it is essential to assess the model's performance using criteria that measure both the model's fit and the accuracy of selected variables. Two commonly used criteria in model selection for factorial experiments are the Expected Error Rate (EER) and the Interaction Error Rate (IER). The Expected Error Rate (EER) is a measure of how well the regression model predicts the response variable, providing an estimate of the average prediction error across all experimental runs [9]. The EER is defined as:

$$EER = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \qquad (2)$$

Where (N) is the total number of observations, (Y_i) is the observed response, ($\hat{Y}_i$) is the predicted response from the model. A lower EER indicates better model performance, as it reflects a smaller discrepancy between observed and predicted values. In hypertension studies, a low EER is critical to ensure the model accurately captures the influence of each factor and interaction on blood pressure levels. The Interaction Error Rate (IER) focuses specifically on the error associated with interaction terms in the model. It measures how well the model captures the effects of interactions between factors [10], which is particularly important in factorial designs where the combined influence of multiple factors is of interest.

IER is calculated as:

$$IER = \frac{1}{M} \sum_{i=1}^{M} (\hat{\beta}_{\text{interaction,j}} - \beta_{\text{interaction,j}})^2 \qquad (3)$$

Where, (M) is the number of interaction terms in the model, ($\hat{\beta}_{\text{interaction,j}}$) is the estimated coefficient for the i^{th} interaction, ($\beta_{\text{interaction,j}}$) is the true coefficient for the i^{th} interaction. A lower IER indicates that the model is effectively capturing the interaction effects, providing a reliable estimate of how combined factors influence blood pressure. In factorial designs, achieving a low IER is essential to ensure the model accurately represents the complex relationships between factors and their interactions.

2. Adaptive Lasso

Adaptive Lasso is a modification of the standard Lasso method, designed to apply different penalties to different coefficients, thereby overcoming some limitations of traditional Lasso [12] . Adaptive Lasso uses a two-step approach: Initial Estimation: First, coefficients are estimated using ordinary least squares (OLS) or another estimator to provide initial values. Weighted Penalty Application: In the second step, Lasso is applied with penalty weights inversely proportional to the initial coefficient estimates, meaning that coefficients with larger initial estimates face smaller penalties. This adaptive weighting allows the method to retain more influential variables by applying less shrinkage to them, which makes it particularly effective in high-dimensional contexts with multicollinearity, as is often seen in factorial experiments [13]. This property is valuable when analyzing hypertension, where complex interactions between lifestyle, genetic, and environmental factors might all contribute to blood pressure levels. The objective function for the adaptive Lasso is [14]:

$$\min\{\sum_{i=1}^{N}(Y_i - X_i^T\beta)^2 + \lambda \sum_{j=1}^{p} w_i\,|\beta_j|\} \qquad (4)$$

Where, (Y_i) represents the response variable (blood pressure levels), $(X_i\ )$ is the vector of predictors for the i^{th} observation, $(\beta\ )$ is the vector of regression coefficients, (w_i) are the adaptive weights, calculated based on initial estimates $(\hat{\beta}_j\ )$, $(\lambda\ )$ is a tuning parameter that controls the strength of the penalty.

Tuning the penalty parameter ,To optimize the adaptive Lasso model, we tune the penalty parameter $(\lambda\ )$ through cross validation [15]. Cross-validation helps us select a value of $(\lambda\ )$ that minimizes the model's prediction error, ensuring that only relevant factors and interactions are retained without overfitting. This tuning process is crucial in identifying the factors most strongly associated with hypertension while excluding less impactful variables.

3. Real Data Analysis

This section examines real data on blood pressure (response variable) and its relationships with key genetic, lifestyle, and environmental factors for sample size (63) visitors. Data were collected from Ibn Sina Teaching Hospital – Mosul, Iraq. Adaptive Lasso was applied to identify the most significant factors and interactions affecting blood pressure. Blood pressure, a critical indicator linked to cardiovascular health, is influenced by genetics, diet, lifestyle, and stress. Chronic high blood pressure increases the risk of severe complications, such as heart attacks and strokes. This study focuses on six primary factors:

Age (X1): Younger (level 1) and older (level 2), with risk increasing with age. Genetic History (X2): No family history (level 1) and family history of hypertension (level 2), affecting susceptibility. Salt Intake (X3): Low (level 1) and high (level 2), where high intake correlates with increased blood pressure. Physical Activity (X4): Active (level 1) and sedentary (level 2), with inactivity raising hypertension risk. Blood Viscosity (X5): Normal (level 1) and high (level 2), where higher viscosity may elevate blood pressure. Stress Level (X6): Low (level 1) and high (level 2), as sustained stress can raise blood pressure. The study examines both individual and interactive effects of these factors to provide a comprehensive understanding of hypertension risks, guiding targeted health interventions.

Normality tests were used to assess if the dataset follows a normal distribution, an essential assumption in statistical analyses. Tests such as the Kolmogorov-Smirnov and Shapiro-Wilk were applied, and visual aids like histograms were used to verify data alignment with the normal distribution, supporting the reliability of subsequent statistical evaluations.

The table below presents the Kolmogorov-Smirnov and Shapiro-Wilk test results for each main factor (X1 to X6), indicating that all factors follow a normal distribution.

Table 1: Kolmogorov-Smirnov and Shapiro-Wilk test results

Main Factors	Kolmogorov-Smirnov	P-Value (KS)	Shapiro-Wilk	P-Value (SW)
x1	0.052	0.76	0.989	0.912
x2	0.045	0.845	0.978	0.823
x3	0.039	0.908	0.981	0.71
x4	0.061	0.625	0.992	0.867
x5	0.05	0.784	0.987	0.936
x6	0.048	0.812	0.985	0.902

Table 1 presents the results of the Kolmogorov-Smirnov (KS) and Shapiro-Wilk (SW) tests for normality across the main factors x1 to x6. For each factor, both tests yield (p) values greater than 0.05, indicating no significant deviation from a normal distribution. This suggests that all main factors (x1) through (x6) follow a normal distribution, which is suitable for further statistical analysis based on the assumption of normality.

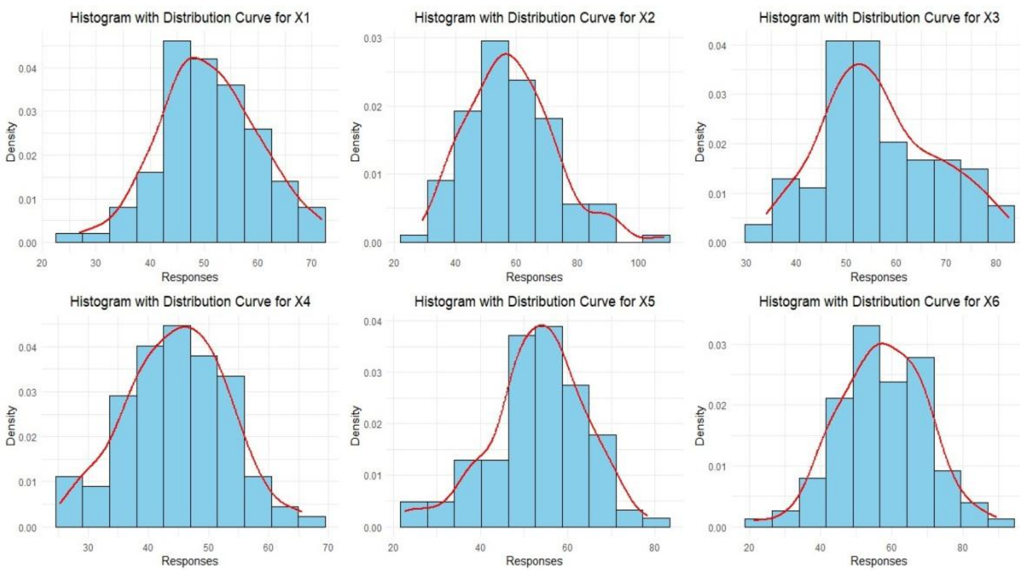

Figure 1 : Normal distribution of major factors

Table 2: Estimated Coefficients for Regression Model

Variable	$\hat{\beta}$	Variable	$\hat{\beta}$	Variable	$\hat{\beta}$	Variable	$\hat{\beta}$
X1	4.512	X3X5	3.208	X2X3X5	0.009	X1X3X4X6	1.543
X2	6.334	X3X6	0.005	X2X3X6	1.231	X1X3X5X6	0.009
X3	5.121	X4X5	0.028	X2X4X5	0.004	X1X4X5X6	2.003
X4	2.018	X4X6	1.143	X2X4X6	0.012	X2X3X4X5	0.009
X5	8.032	X5X6	0.005	X2X5X6	2.223	X2X3X4X6	1.623
X6	1.903	X1X2X3	8.201	X3X4X5	0.003	X2X3X5X6	0.004
X1X2	10.101	X1X2X4	0.007	X3X4X6	2.001	X2X4X5X6	0.008
X1X3	2.488	X1X2X5	7.648	X3X5X6	0.002	X3X4X5X6	1.002
X1X4	1.803	X1X2X6	0.004	X4X5X6	0.000	X1X2X3X4X5	1.89
X1X5	6.894	X1X3X4	0.289	X1X2X3X4	1.234	X1X2X3X4X6	0.005
X1X6	0.021	X1X3X5	2.341	X1X2X3X5	0.003	X1X2X3X5X6	0.001
X2X3	2.312	X1X3X6	0.002	X1X2X3X6	0.007	X1X2X4X5X6	0.002
X2X4	0.003	X1X4X5	0.003	X1X2X4X5	0.001	X1X3X4X5X6	0.000
X2X5	1.657	X1X4X6	2.002	X1X2X4X6	0.002	X2X3X4X5X6	0.007
X2X6	0.012	X1X5X6	0.001	X1X2X5X6	1.201	X1X2X3X4X5	1.973
X3X4	0.007	X2X3X4	3.002	X1X3X4X5	0.000		

Where the values of EER and IER are (0.09, 0.002) respectively and that indicates the well performance of the proposed model.

5. Conclusions

This study combined factorial analysis with Adaptive Lasso to identify and evaluate the significant factors and interactions that influence blood pressure. By examining six primary factors age (X1), genetic history (X2), salt intake (X3), physical activity (X4), blood viscosity (X5), and stress level (X6) the study aimed to provide a comprehensive understanding of how these elements impact hypertension, both individually and in combination. The results reveal that age and genetic history play critical roles as individual predictors of hypertension risk, aligning with established medical knowledge. Additionally, salt intake and physical activity were identified as modifiable lifestyle factors substantially influencing blood pressure levels. Notably, the analysis of interactions provided insights into how these factors compound each other's effects. For example, the combined impact of high salt intake and stress level was observed to significantly elevate blood pressure beyond their individual effects, highlighting the complexity of managing hypertension risk. The use of Adaptive Lasso was instrumental in refining the model by selecting only the most impactful variables, which enhanced model interpretation ability and accuracy. This selective approach allowed us to address the high dimensional nature of factorial designs effectively, reducing potential noise from less influential interactions and focusing on the key drivers of blood pressure.

In conclusion, the findings of this study emphasize the multifactorial nature of hypertension, with both lifestyle and genetic factors contributing to its onset and severity. These results underscore the importance of targeted intervention strategies that consider multiple risk factors and their interactions. Future research could expand on this study by including additional health and environmental variables or exploring nonlinear interactions, thereby deepening our understanding of hypertension and informing more personalized health recommendations.

References

[1] Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo Jr, J. L., ... & National High Blood Pressure Education Program Coordinating Committee. (2003). Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. hypertension, 42(6), 12061252.

[2] Draper, N. R. (1998). Applied regression analysis. McGrawHill. Inc.

[3] EP, B. G. (1978). Statistics for Experimenters: An Introduction to Design. Data Analysis, and Model Building, 4348.

[4] Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction.

[5] Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1), 1.

[6] Meinshausen, N., & Bühlmann, P. (2010). Stability selection. Journal of the Royal Statistical Society Series B: Statistical Methodology, 72(4), 417473.

[7] Montgomery, D. C. (2017). Design and analysis of experiments. John wiley & sons.

[8] Myers, R. H., Montgomery, D. C., & AndersonCook, C. M. (2016). Response surface methodology: process and product optimization using designed experiments. John Wiley & Sons.

[9] Neter, J., Wasserman, W., & Kutner, M. H. (1983). Applied linear regression models. Richard D. Irwin.

[10] Soliman, E. Z., Ambrosius, W. T., Cushman, W. C., Zhang, Z. M., Bates, J. T., Neyra, J. A., ... & Lewis, C. E. (2017). Effect of intensive blood pressure lowering on left ventricular hypertrophy in patients with hypertension: SPRINT (Systolic Blood Pressure Intervention Trial). Circulation, 136(5), 440450.

[11] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267288.

[12] Wu, C. J., & Hamada, M. S. (2011). Experiments: planning, analysis, and optimization. John Wiley & Sons.

[13] Wu, C. J., & Hamada, M. S. (2011). Experiments: planning, analysis, and optimization. John Wiley & Sons.

[14] Wu, T. T., Chen, Y. F., Hastie, T., Sobel, E., & Lange, K. (2009). Genomewide association analysis by lasso penalized logistic regression. Bioinformatics, 25(6), 714721.

[15] Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 14181429.

Zou, H., & Li, R. (2008). One step sparse estimates in nonconcave penalized likelihood models. Annals of statistics, 36(4), 1509