The risk factors of resting blood pressure for shock patients

Shayamsundar Karmakar

Dept . Of statistics , university of burdwan

ABSRACT

Of the Project Work

- Elevated blood pressure is a fundamental cardiovascular risk factor. Systolic and diastolic blood pressure are mainly considered in hypertension management. The current report considers the role of BP for some heart patients. It is found that mean BP is higher at older ages ($P < 0.0001$). It is higher for diabetes heart patients ($P = 0.0836$), and it increases as the maximum heart rate ($P = 0.0092$) rises, or ST depression induced by exercise relative to rest ($P = 0.01366$) increases. It is higher for the patients having resting electrocardiographic result at normal ($P=0.0388$), or thalasemia at normal level ($P = 0.1189$) than others. Variance of BP is higher at older ages ($P=0.0091$), or patients without chest pain ($P = 0.1113$).
Keywords: Role of Blood; Heart Patients .

Objectives

…………………………

1: What is the relationship RestingBP with the rest of the others variablis/factors.

2: What are the risk factor for restingBP.

3: What are the protecting factor for RestingBP.

Multiple Regression

....

- **Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regressions (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable.**
- **We generally use the following formula to fit a 'Multiple Linear**
- **Regression Model':** $yi = \beta 0 + \beta 1xi1 + \beta 2xi2 + \cdots + \beta pxip + \epsilon i$ $, i = 1,2, \ldots, n$
- For p independent variables where,
- $yi = i^{th}$ value of the dependent variable,
- $xij = i^{th}$ value of the j^{th} independent variable, $j = 1,2, \ldots, p$
- β_0 = y-intercept (Constant term)
- βj = regression (slope) coefficient for the j^{th} independent variable, $j = 1,2, \ldots, p$
- ϵi = error term (also called residuals) associated with the i^{th} predicted value.

Assumptions

- Linear regression makes several assumptions about the data, such as:

1. Linearity of the data. The relationship between the predictor (x) and the outcome (y) isassumed to be linear.
2. Normality of residuals. The residual errors are assumed to be normally distributed.
3. Homogeneity of residuals variance. The residuals are assumed to have a constant variance (homoscedasticity).
4. Independence of residuals error terms.

- We should check whether or not these assumptions hold true. Potential problems include:

1. Non-linearity of the outcome - predictor relationships
2. Heteroscedasticity: Non-constant variance of error terms.
3. Presence of influential values in the data that can be:
   - Outliers: extreme values in the outcome (y) variable
   - High-leverage points: extreme values in the predictors (x) variable

- All these assumptions and potential problems can be checked by producing somediagnostic plots visualizing the residual errors.

Regression Diagnostics Plots

……………………………………………………

- The diagnostic plots show residuals in four different ways:
-

1. Residuals vs Fitted. Used to check the linear relationship assumptions. A horizontal line, without distinct patterns is an indication for a linear relationship, what is good for the model.
2. Normal Q-Q. Used to examine whether the residuals are normally distributed. It's good if

- residuals points follow the straight dashed line.
- Scale-Location (or Spread-Location). Used to check the homogeneity of variance of the residuals (homoscedasticity). Horizontal line with equally spread points is a good indication of homoscedasticity. This is not the case in our example, where we have a heteroscedasticity problem

Analysis

…………………………..

Residuals:

Min	1Q	Median	3Q	Max
-47.796	-10.974	-1.700	9.101	65

Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	113.28521 4	8.029247	14.109	< 2e-16 ***
Age	0.449337	0.076274	5.891	5.9e-09 ***
Sex0	-0.037419	1.660676	-0.023	0.982030
ChestPainType2	-0.708789	3.447608	-0.206	0.837171
ChestPainType3	-1.987489	3.227836	-0.616	0.538266
Cholesterol	0.020773	0.006179	3.362	0.000815 ***
FastingBSfalse	-0.833611	1.636447	-0.509	0.610627
RestingECG1	2.754274	1.496429	1.841	0.066100 .
MaxHR	-0.036990	0.029930	-1.236	0.216901

Coefficients

……………………………………..

ExerciseAngina0	-3.760485	1.660725	-2.264	0.023850 *
Oldpeak	1.433790	0.755026	1.899	0.057968 .
ST_Slope2	-0.052208	1.879188	-0.028	0.977844
ST_Slope3	-10.217406	3.067013	-3.331	0.000909 ***
HeartDisease	-1.024476	2.015927	-0.508	0.611476

The initial multiple regression model fitted to the data is asfollows:

………………………………………………………………….

- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
- Residual standard error: 16.82 on 715 degrees of freedom (188 observations deleted due to missingness)
- Multiple R-squared: 0.1326, Adjusted R-squared: 0.1168
- F-statistic: 8.407 on 13 and 715 DF, p-value: 5.493e-16
- ŷ= 113.285214 + 0.449337 x1 -0.037419 x2 - 0.708789 x3 -1.987489 x4 + 0.
- 020773 x5 -0.833611 x6 +2.754274x7 -0.036990 x8 -3.760485 x9 + 1.433790
- x10-0.052208 x11-10.217406 x12 -1.024476 x13

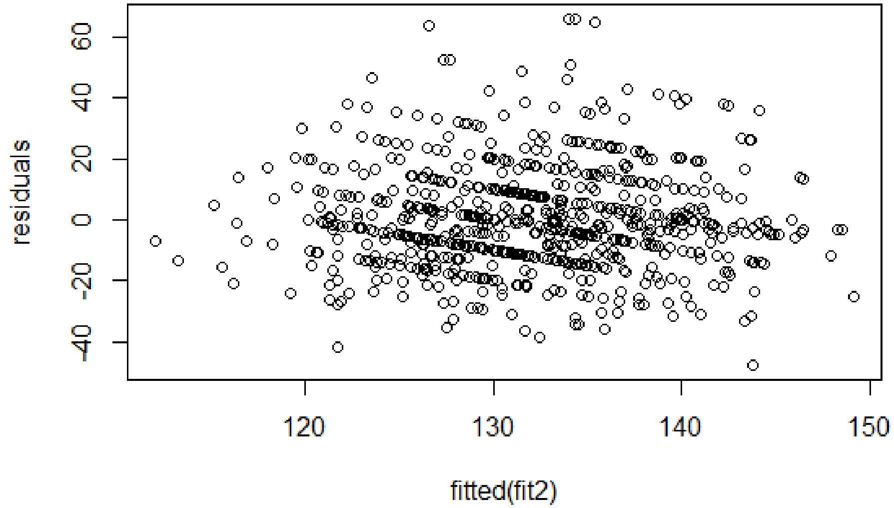
VIF Values as obtained from R :

Multicollinearity Diagnostics in The Model:

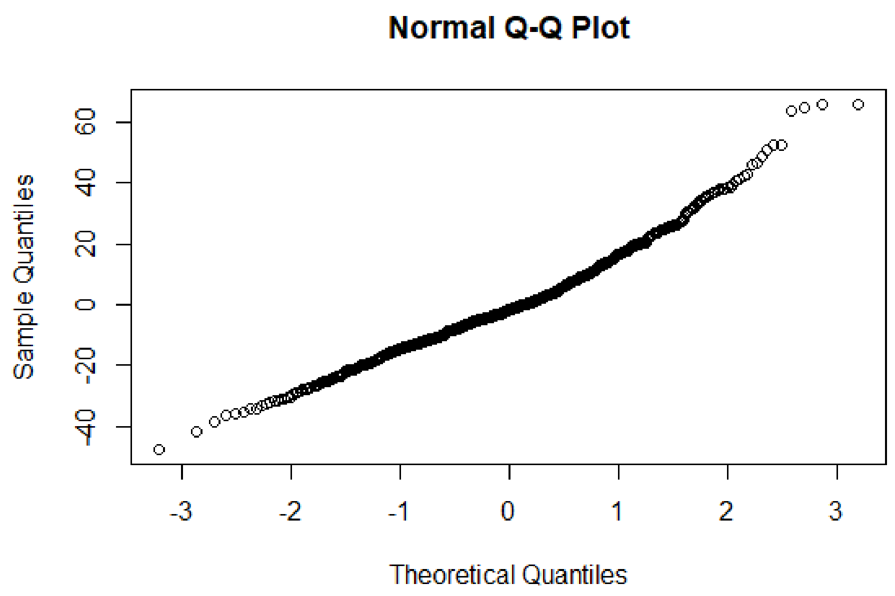
…………………………………………………………………….

	GVIF	Df	GVIF^(1/(2*Df))
Age	1.340834	1	1.157944
Sex	1.138512	1	1.067011
ChestPainType	1.372950	2	1.082465
Cholesterol	1.254460	1	1.120027
FastingBS	1.244326	1	1.115493
RestingECG	1.065197	1	1.032084
MaxHR	1.508870	1	1.228361
ExerciseAngina	1.718047	1	1.310743
Oldpeak	1.567956	1	1.252181
ST_Slope	2.364070	2	1.239981
HeartDisease	2.592471	1	1.610115

Residuals Plot



Normal Q-Q Plot



“ BOX-COX Transformation

A Box Cox transformation is a way to transform non-normal dependent variables into a normal shape. Normality is an important assumption for many statistical techniques;
if our data isn’t normal, applying a Box-Cox means that we are able to run a broader number of tests.
The Box Cox transformation is named after statisticians George Box and Sir David Roxbee Cox who collaborated on a 1964 paper and developed the technique.
At the core of the Box Cox transformation is an exponent, lambda (λ), which varies from -5 to 5. All values of λ are considered and the optimal value for our data is selected;
The “optimal value” is the one which results in the best approximation of a normal distribution curve. The transformation of Y has the form:

$$y(\lambda) = \frac{y^{\lambda}-1}{\lambda} \text{ if } \lambda \neq 0,$$
$$\log(y) \text{ if } \lambda = 0$$

The test works only for positive data.
There is an extreme situation, called multicollinearity, where collinearity exists between three or more variables even if no pair of variables has a particularly high correlation.
This means that there is redundancy between predictor variables.
In the presence of multicollinearity, the solution of the regression model becomes unstable.

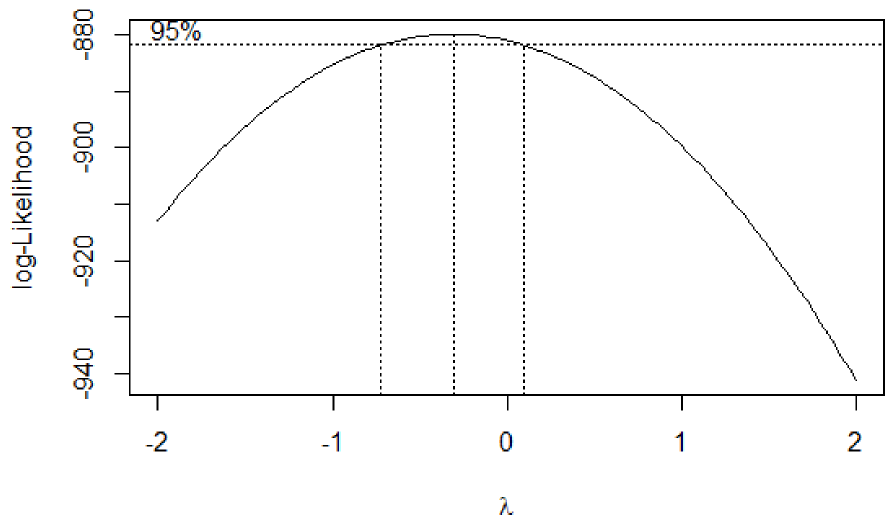
- Model Fitting After Box-Cox Transformation

We have performed the ‘Box-Cox’ transformation to the response variable BCF. From the curve of the log-likelihood function we have determined the optimal value of lambda viz.,
$\boldsymbol{\lambda_{opt} = -0.3030303}$

and the corresponding value of the log-likelihood function is −3051.478. Here we denote the fitted values by $\boldsymbol{y^{bc}}$. We are presenting below the graph of the log-likelihood function.

$\boldsymbol{Prime}$

Log-likelihood of Box Cox Transformation



Residuals

Min	1Q	Median	3Q	Max
-0.101614	-0.017760	-0.000993	0.017056	0.090412

Coefficients

	Estimate Std.	Error	t value	Pr(>|t|)
(Intercept)	2.510e+00	1.363e-02	184.161	< 2e-16 ***
Age	8.090e-04	1.295e-04	6.249	7.1e-10 ***
Sex0	-1.149e-03	2.819e-03	-0.408	0.683671
ChestPainType2	-1.121e-03	5.852e-03	-0.192	0.848144
ChestPainType3	-3.427e-03	5.479e-03	-0.626	0.531791
Cholesterol	3.892e-05	1.049e-05	3.712	0.000222 ***
FastingBSfalse	-8.204e-04	2.778e-03	-0.295	0.767803
RestingECG1	4.668e-03	2.540e-03	1.838	0.066504 .
MaxHR	-5.715e-05	5.080e-05	-1.125	0.260971
ExerciseAngina	0 -6.083e-03	2.819e-03	-2.158	0.031260 *
Oldpeak	2.689e-03	1.282e-03	2.098	0.036260 *
ST_Slope2	1.346e-04	3.190e-03	0.042	0.966342
ST_Slope3	-1.875e-02	5.206e-03	-3.601	0.000339 ***
HeartDisease	-2.148e-03	3.422e-03	-0.628	0.530445

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’	0.1 ‘ ’	1

Residual standard error: 0.02854 on 715 degrees of

freedom (188 observations deleted due to missingness)

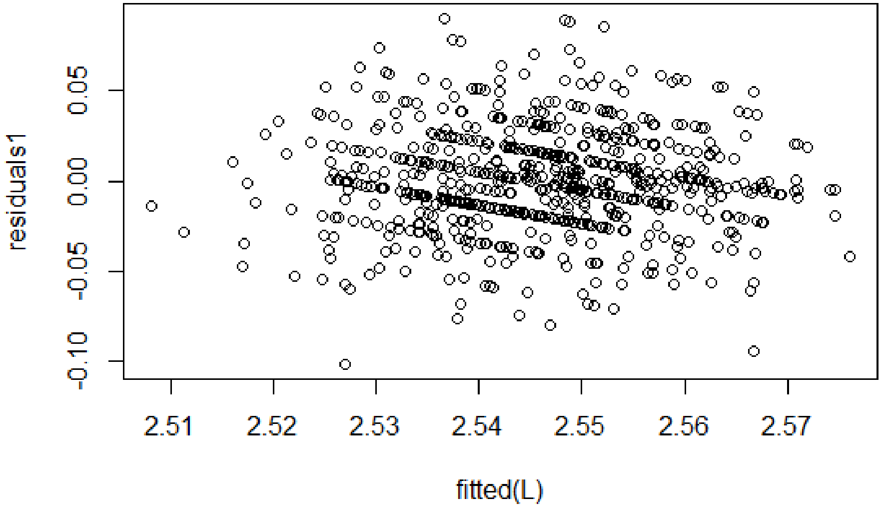
Multiple R-squared: 0.1429, Adjusted R-squared: 0.1273

F-statistic: 9.17 on 13 and 715 DF, p-value: < 2.2e-16

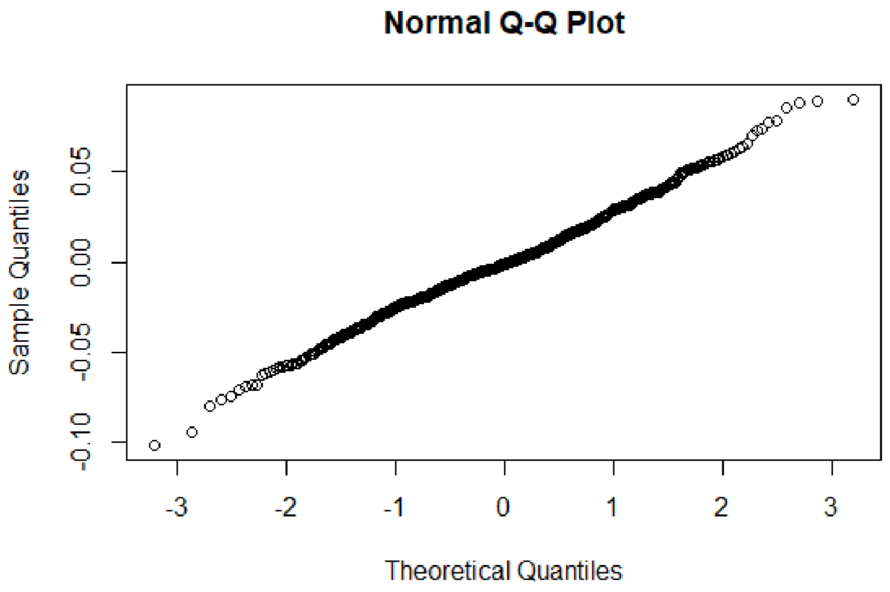
$$\hat{\mathbf{y}} = 2.510\text{e+}00 + 8.090\text{e-}04\ x_1 - 1.149\text{e-}03 x_2 - 1.121\text{e-}03\ x_3 - 3.427\text{e-}03 x_4 + 3.892\text{e-}05 x_5 - 8.204\text{e-}\ 04 x_6$$

$$+ 4.668\text{e-}03\ x_7 - 5.715\text{e-}05\ x_8 - 6.083\text{e-}03\ x_9 + 2.689\text{e-}03\ x_{10} + 1.346\text{e} - 04\ x_{11} - 1.875\text{e-}02 x_{12} - 2.148\text{e}03 x_{13}$$

Residuals Plot



Normal Q-Q Plot



❖ Fitting Of Generalized Linear Model To The Response Variable:

We have fitted a 'Generalized Linear Model' to our data. Since the response variable viz., BCF is positive valued and continuous, we assume them to belong to the "gamma" family and take a "log" link function. Before going to the detailed analyses of the fitted model we shall give a brief description of the terminologies concerned to a 'Generalized Linear Model'.

▪ Generalized Linear Model:

In a Generalized Linear Model (GLM)

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i$$

the response y_i , i=1, 2, ... ,n is modeled by a linear function of the explanatory variables x_j , j=1, 2, ... , p plus an error term. That is a GLM is made up of a linear predictor of the form

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

And two functions,

- A link function that describes how the mean $\mu_i = E(y_i)$ depends on the linear predictor

$$g(\mu_i) = \eta_i$$

- A variance function that describes how the variance, $var(y_i)$ depends on the mean $var(y_i) = \Phi V(\mu)$ where the dispersion parameter Φ is a constant.

Note that in a GLM the response variable is assumed to belong to the 'Exponential Family' of distributions.

Residual:

Min	1Q	Median	3Q	Max
-47.796	-10.974	-1.700	9.101	65.965

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	113.285214	8.029247	14.109	< 2e-16 ***
Age	0.449337	0.076274	5.891	5.9e-09 ***
Sex0	-0.037419	1.660676	-0.023	0.982030
ChestPainType2	-0.708789	3.447608	-0.206	0.837171
ChestPainType3	-1.987489	3.227836	-0.616	0.538266
Cholesterol	0.020773	0.006179	3.362	0.000815 ***
FastingBSfalse	-0.833611	1.636447	-0.509	0.610627
RestingECG1	2.754274	1.496429	1.841	0.066100 .
MaxHR	-0.036990	0.029930	-1.236	0.216901
ExerciseAngina	0 -3.760485	1.660725	-2.264	0.023850 *
Oldpeak	1.433790	0.755026	1.899	0.057968 .
ST_Slope2	-0.052208	1.879188	-0.028	0.977844
ST_Slope3	-10.217406	3.067013	-3.331	0.000909 ***
HeartDisease	-1.024476	2.015927	-0.508	0.611476

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

GLM fitting in R:

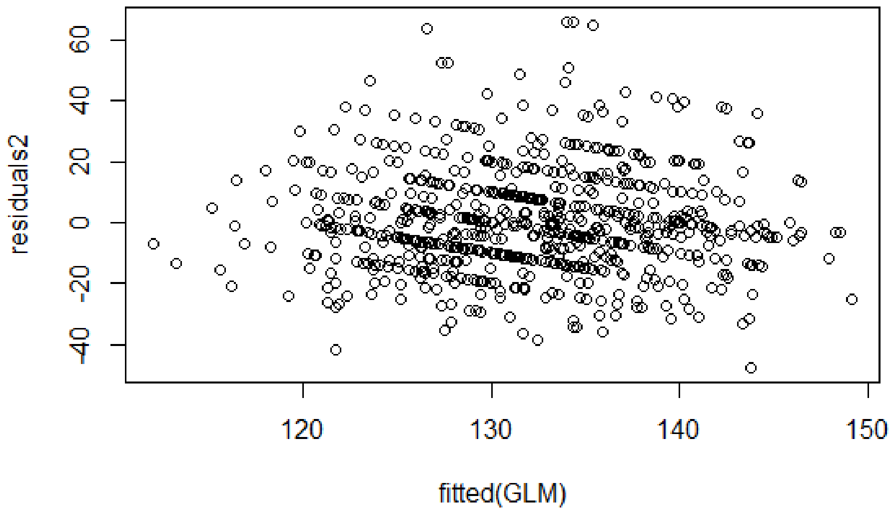
$\hat{\mathbf{y}} =$ 113.285214 + 0.449337 x_1 -0.037419 x_2 -0.708789 x_3 -1.987489 x_4 + 0.020773 x_5 -0.833611 x_6 +2.754274x_7 -0.036990 x_8 -3.760485 x_9 + 1.433790 x_{10}-0.052208 x_{11}-10.217406 x_{12} -1.024476 x_{13}

(Dispersion parameter for gaussian family taken to be 282.8305)

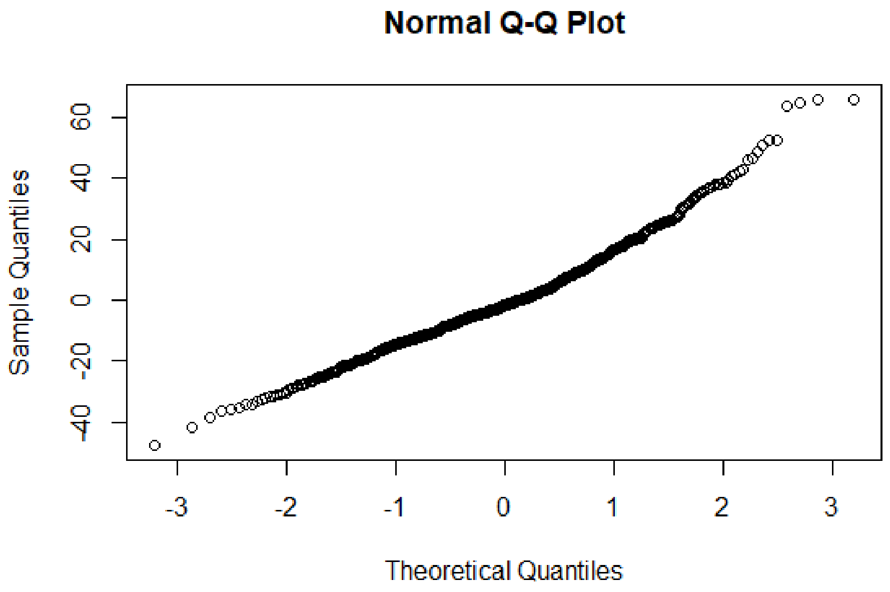
Null deviance: 233135 on 728 degrees of freedom Residual deviance: 202224

on 715 degrees of freedom (188 observations deleted due to missingness)
AIC: 6199.8
Number of Fisher Scoring iterations: 2

Residuals plot (GLM)



Normal Q-Q Plot



Fitting Of Joint Generalized Linear Model:

Before going to the detailed discussions and results about BCF we shall be concerned about the general structures of the "Joint Generalized Linear Model" (JGLM).

- In GLM we have only the model for assumming the variance is constant.If variance is not constant , we have to consider mean and variance modeling jointly this processes is called joint GLM.

Statistical methods

The data set is a multivariate form, and the response resting blood pressure is heterogeneous, positive and continuous, which can be modeled by using appropriate transformation if the variance is stabilized under that transformation. But resting BP is not stabilized. Therefore, it should be modeled by joint generalized linear models (JGLMs). Thus, we have examined resting BP using JGLMs under both the Log-normal and Gamma distributions, which are clearly given in the book by Lee., *et al* [12]. It is also well discussed in many research papers by Lesperance and Park [13], Das and Lee [14], Qu., *et al.* [15], which are not reported herein. For details on JGLMs discussions, readers can see the book by Lee., *et al* [12].

<u>Statistical and graphical analysis:</u> The response resting BP is treated as the dependent variable and the remaining other factors and variables are treated as the independent variables. Herein resting BP has been modeled by JGLMs under both the Gamma and Log-normal distributions. The final model has been accepted considering the lowest Akaike information criterion (AIC) value (within each class), which minimizes both the squared error loss and predicted additive errors. AIC rule is clearly described in the book by Hastie., et al [16]. All the included effects in both the models are not significant. The resting BP analysis results are displayed in Table 1. According to AIC rule (Table 1), Log-normal fit (AIC = 2518) gives better than the Gamma fit (AIC = 2520.157).

Model	Factors	Log-normal model fit				Gamma model fit			
		estimate	s.e.	t-value	P- value	Estimate	s.e.	t-value	P- value
Mean	Constant	4.5199	0.08574	52.717	<0.0001	4.5123	0.08600	52.471	<0.0001
	Age	0.0037	0.00085	4.362	<0.0001	0.0039	0.00085	4.557	<0.0001
	Cholestoral	0.0002	0.00014	1.141	0.2547	0.0002	0.00014	1.211	0.2269
	Thalach	0.0009	0.00036	2.622	0.0092	0.0010	0.00036	2.677	0.0078
	Oldpeak	0.0166	0.00670	2.481	0.0137	0.0164	0.00672	2.437	0.0154
	Fbs	0.0352	0.02029	1.736	0.0836	0.0356	0.02032	1.751	0.0809
	Restecg	-0.0289	0.01392	-2.075	0.0389	-0.0285	0.01396	-2.044	0.0418
	Thal 2	-0.0474	0.03030	-1.564	0.1189	-0.0492	0.03039	-1.618	0.1067
	Thal 3	-0.0321	0.03011	-1.068	0.2864	-0.0316	0.03020	-1.047	0.2959
Dispersion	Constant	-5.521	0.5348	-10.323	<0.0001	-5.509	0.5365	-10.269	<0.0001
	Age	0.025	0.0094	2.625	0.0091	0.025	0.0094	2.606	0.0096
	Chest pain 2	-0.398	0.2491	-1.597	0.1113	-0.387	0.2489	-1.555	0.1.210
	Chest pain 3	0.014	0.1836	0.079	0.9371	-0.001	0.1835	-0.008	0.9936
AIC		2518				2520.157			

***Table 1:** Results for mean and dispersion models for resting BP from Gamma and Log-Normal fit.*

Absolute Residuals Plot & Normal Plot

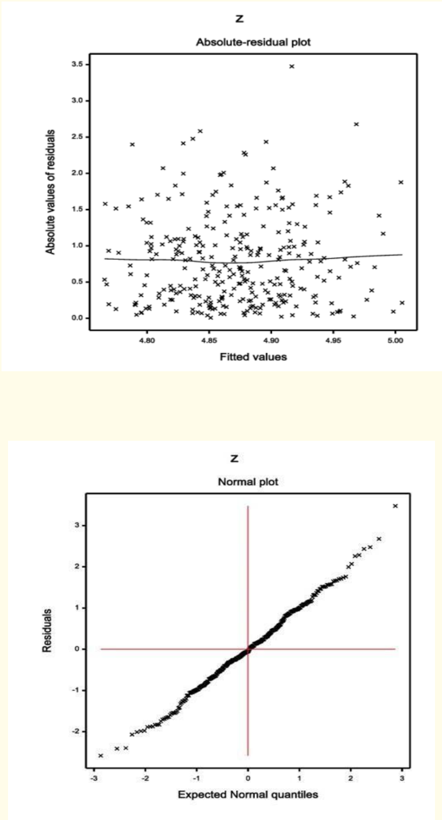


Figure 1: *For the joint Log-normal fitted resting BP model (Table 1), the (a) absolute residuals plot with respect to the fitted values, and (b) the normal probability plot for the mean model.*

<u>Results</u>: Summarized JGLMs results for resting BP analysis are displayed in Table 1. Log-normal fitted resting BP model (Table 1) indicates that mean BP is higher at older ages ($P < 0.0001$). It is higher for diabetes heart patients ($P = 0.0836$), and it increases as the maximum heart rate ($P = 0.0092$) rises, or ST depression induced by exercise relative to rest ($P = 0.01366$) increases. It is higher for the patients having resting electrocardiographic result at normal ($P = 0.0388$), or thalasemia at normal level ($P = 0.1189$) than others. Variance of BP is higher at older ages ($P = 0.0091$), or patients without chest pain ($P = 0.1113$).

Log-normal fitted resting BP mean (μ) model (from Table 1) is

$\hat{Z}$ = Log (Resting BP) = 4.5199 + 0.0037 Age + 0.0002 Chol + 0.0009 Thalach + 0.0166 Oldpeak + 0.0352 Fbs – 0.0289 Restecg - 0.0474 (Thal 2) - 0.0321 (Thal 3),

and the Log-normal fitted resting BP variance (σ^2) model is

$\hat{\sigma}^2$ =exp. (-5.521 + 0.025 Age – 0.398 CP(level 2) - 0.014 CP(level 3)).

The mean and variance of resting BP models are given above by two equations.

It is noted that mean resting BP is expressed by Age, Cholestoral, Thalach, Oldpeak, Fbs, Restingecg, Thal while its variance is presented by Age and Chest pain only.

Conclusions: Both the Log-normal and Gamma JGLM fits for resting BP have very similar interpretations (Table 1). The standard errors of the estimators in both the models are very small, concluding that estimates are stable. Graphical diagnostic tools and AIC value are used to accept the appropriate model. The current research should have higher faith in these results than those emanating from Log-Gaussian Citation: Rabindra Nath Das., et al. “Role of Blood Pressure on Heart Patients”. EC Cardiology 7.3 (2020): 01-06. Role of Blood Pressure on Heart Patients 05 (with constant variance), Logistic regression, multiple linear regression, as the current obtained models have been verified by AIC values, graphical diagnostic tools and comparison between two distributions such as Log-normal and Gamma. The current report presents many new interesting results to the cardiology literature. Medical experts and heart patients will be benefitted from the present outputs. Resting BP for heart patients should be examined regularly at older ages, and diabetic and thalassemic heart patients.

THANK YOU