

The relationship between Smoking and Blood Pressure: An analysis on Blood Pressure Prediction

Yiwen Li

Introduction:

- Nearly half of American adults have hypertension, and about 20% of them are unaware of their condition.
- Hypertension can lead to many health problems, including heart attacks, strokes, and heart failure.
- Previous studies have shown that smoking, obesity, sleeping deprivation, and insomnia can increase the risk of hypertension.

Goals of the analysis

- This study aims to examine the relationship between smoking and blood pressure among the American population and to find out what variables are good in predicting blood pressure.
- Some recommendations for the prevention and treatment of hypertension would be provided based on the results at the end of the analysis.

About the data

- This study used the 2011-2012 survey data collected by the National Center for Health Statistics (NCHS) in America. The target population is the non-institutionalized civilian resident population of the United States.
- Our sample data focus on population aged above 17.
- There are 743 observations in the sample data, 500 of which were used to build the model and the rest were used as the testing data.

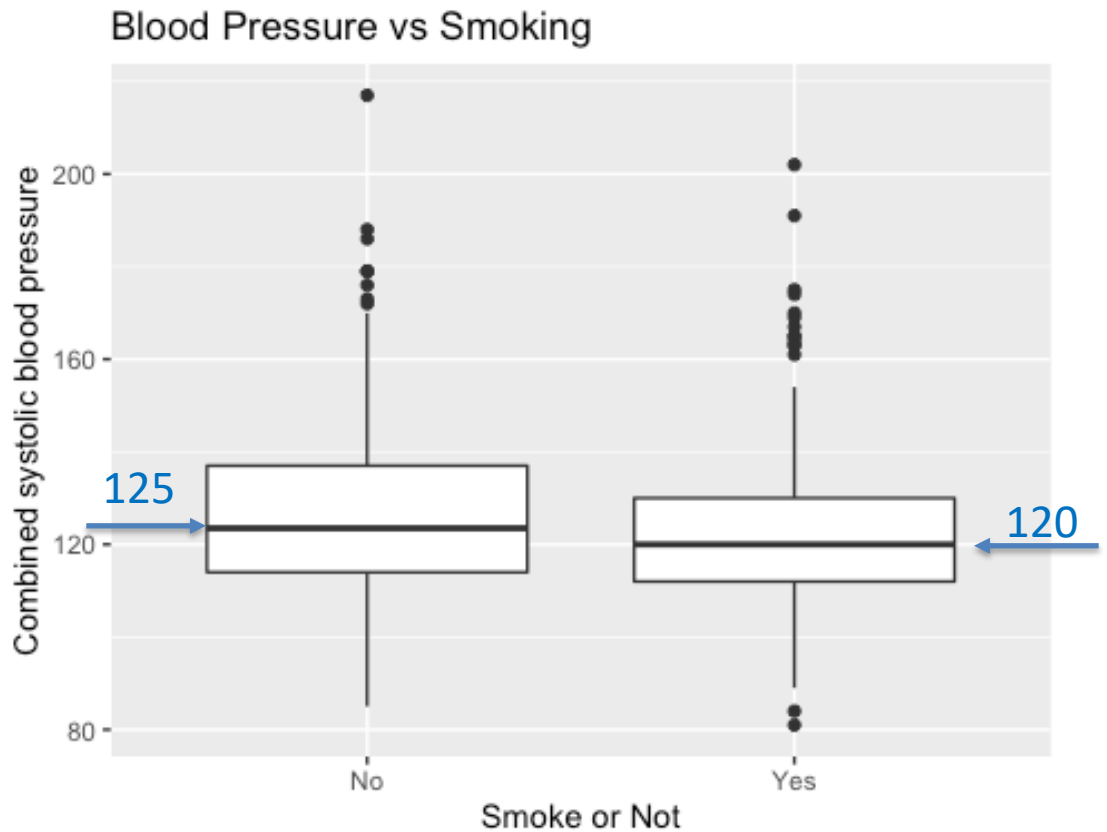
Variables

Numerical variables	Descriptions
Combined systolic blood pressure(BPSysAve)	Measures the pressure in your arteries when your heart beats. <u>Normal systolic pressure should be under 120 mm Hg.</u>
Sleeping Hours per night(SleepHrsNight)	Self-reported number of hours participant usually gets at night on weekdays or workdays.
Body mass index(BMI)	BMI = weight/height ² in kg/m ² . BMI between 18.5 and 25 are indicated as Normal.
Poverty	A ratio of family income to poverty guidelines. <u>Smaller numbers indicate more poverty.</u>

Variables

Categorical variables	Descriptions
Gender	Female or Male
Smoking(SmokeNow)	Yes or No (Answered Yes if participants have smoked 100 or more cigarettes in their lifetime.)
Sleep trouble(SleepTrouble)	Yes or No (Answered Yes if the participant has told a doctor or other health professional that they had trouble sleeping.)

Smoking
vs
Systolic
blood
pressure



Predicting Blood Pressure

Model 1:

$$\widehat{BPSysAve} = \widehat{\beta}_0 + \widehat{\beta}_1 * SmokeNow + \widehat{\beta}_2 * Age + \widehat{\beta}_3 * BMI + \dots$$

Predictors: All variables in the sample data other than ID and Blood pressure

Model 2:

$$\widehat{BPSysAve} = \widehat{\beta}_0 + \widehat{\beta}_1 * SmokeNow + \widehat{\beta}_2 * Gender + \widehat{\beta}_3 * Age + \widehat{\beta}_4 * BMI + \widehat{\beta}_5 * SleepHrsNight$$

Predictors: Smoking, Gender, Age, BMI, Sleeping Hours

VIF of Model 1

High correlation between BMI, Weight, and Height



BMI is a function of Weight and Height, so I removed Weight and Height from the model.

VIF of Model 2

All the VIFs are reasonable, indicating that there are no multicollinearity issues.



No adjustments on Model 2.

Stepwise selection and Lasso shrinkage

Model 1(Stepwise selection using AIC):

Predictors: Gender, Age, Poverty, BMI, SleepTrouble

Model 2(Selected based on previous studies):

Predictors: Gender, Age, BMI, SleepHrsNight, SmokeNow

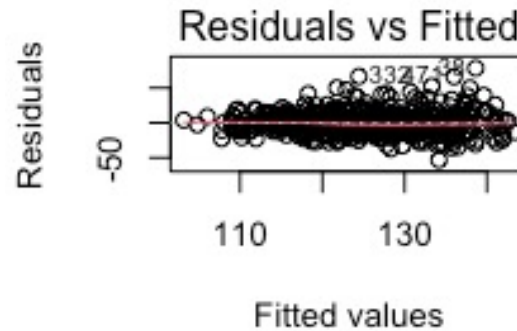
Model 3(Lasso selection):

Predictors: Age

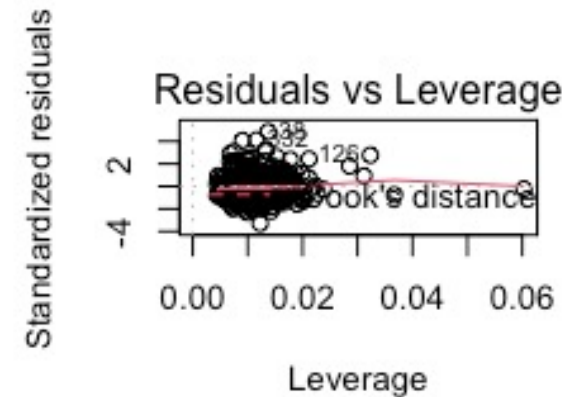
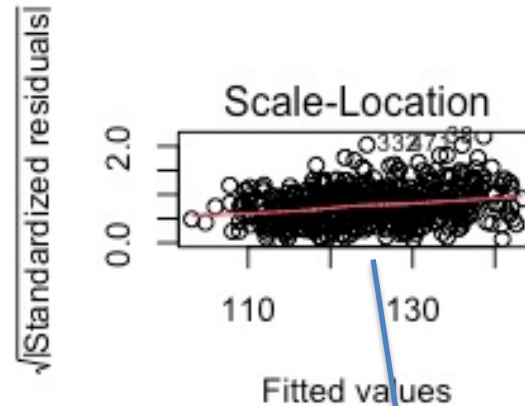
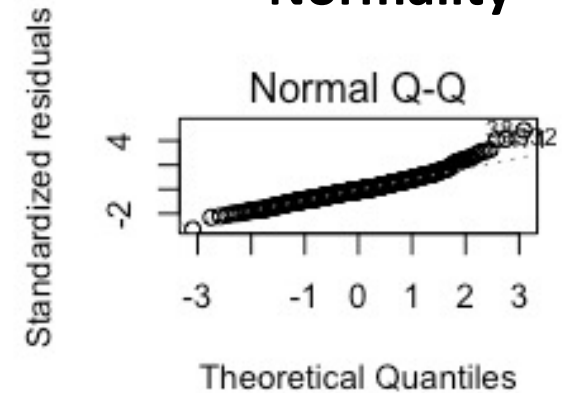
Model Diagnostics

Model 1

✓ **Linearity**



✓ **Normality**

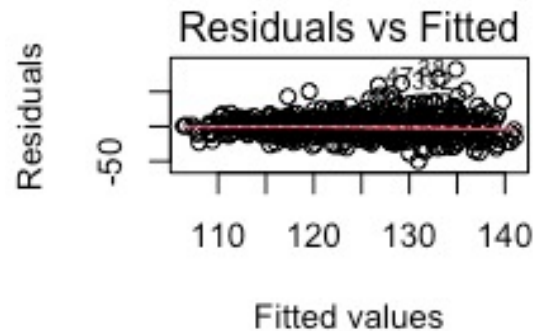


Variance could be improved

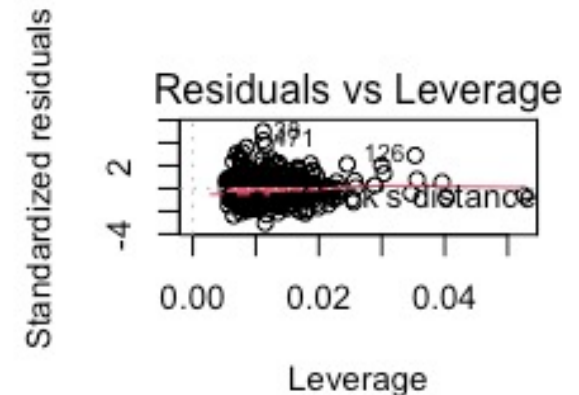
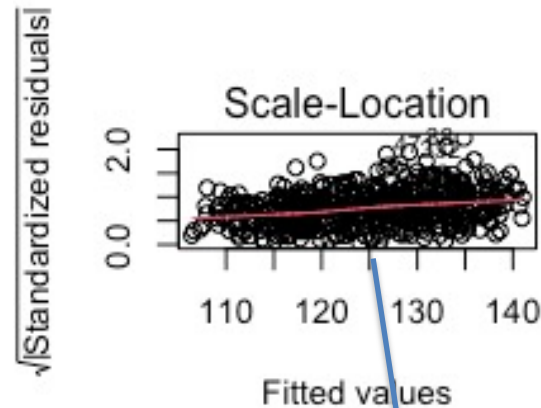
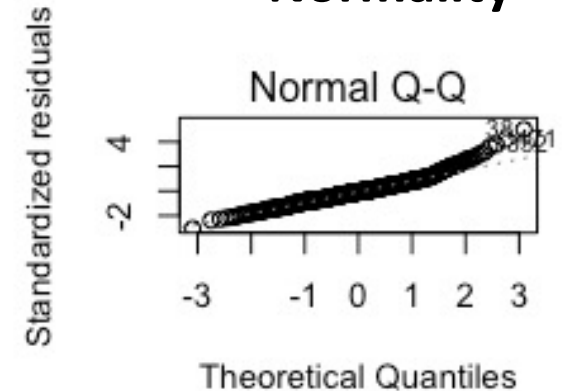
Model Diagnostics

Model 2

✓ **Linearity**



✓ **Normality**

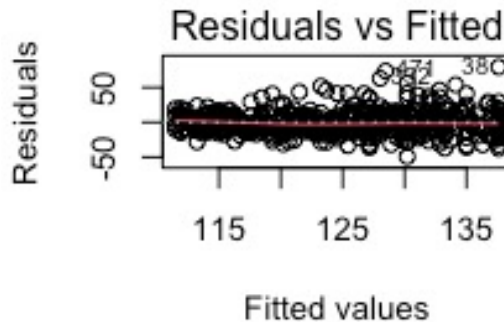


Variance could be improved

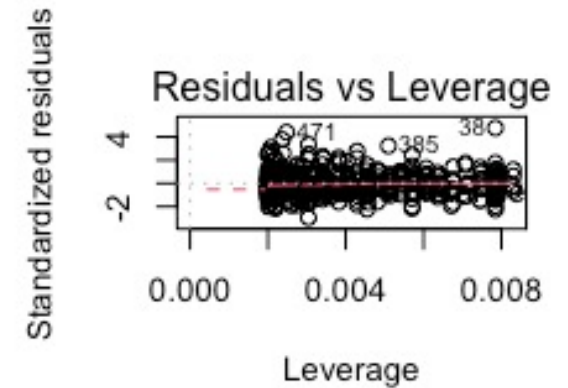
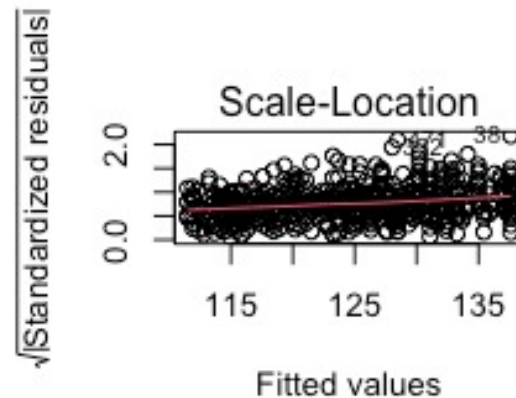
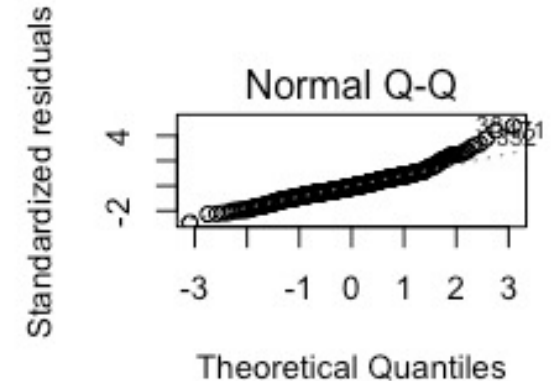
Model Diagnostics

Model 3

✓ **Linearity**



✓ **Normality**



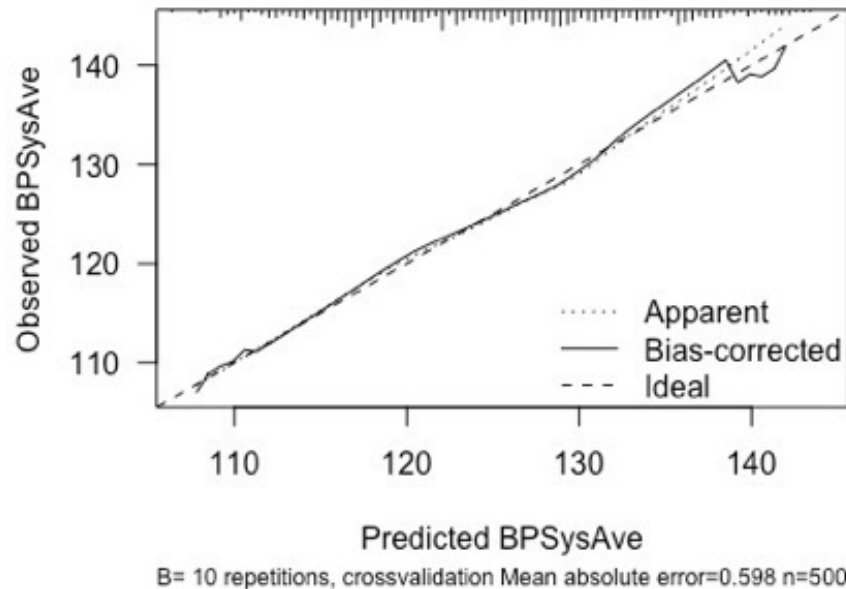
✓ **Constant Variance**

Outliers, Leverages and Influential points

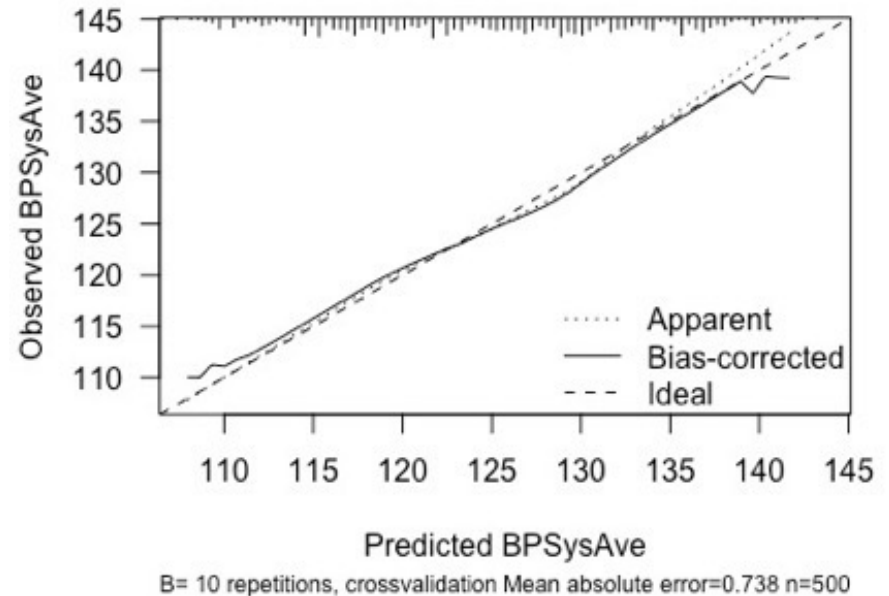
- By calculating the cook's distance, no influential points were found in all three models, but there were outliers and high leverage points.
- I manually checked the sample data and did not find any abnormalities. Therefore, there were no changes made to the data.

Model Validation

Cross-Validation calibration with AIC



Cross-Validation after adding back SmokeNow

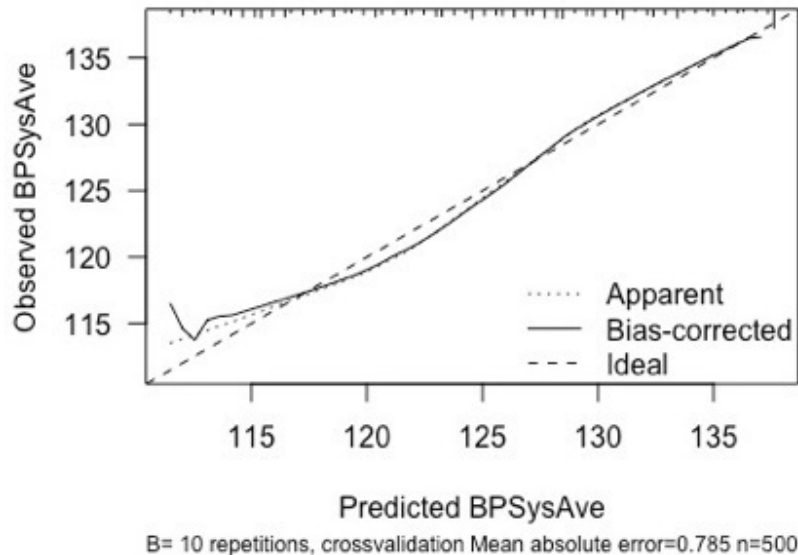


Mean Prediction Error		Mean Prediction Error	
227.951	✓	229.4282	✗

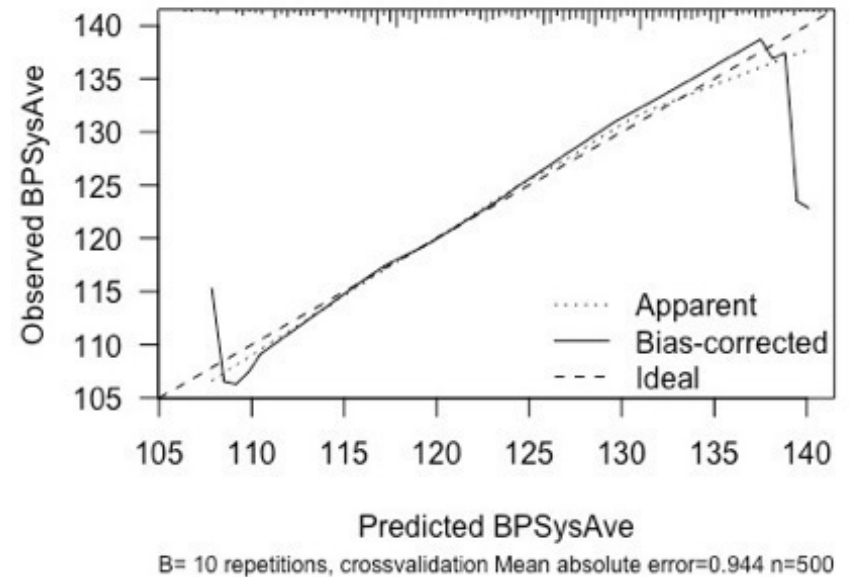
➤ Not include Smoking as a predictor

Model Validation

Cross-Validation calibration with LASSO



Cross-Validation of Model 2



Mean Prediction Error

216.223



Mean Prediction Error

221.274



➤ Drop Model 2 since it has the worst prediction capability

AIC, BIC, and R^2_{adjusted}

Model	AIC	BIC	R^2_{adjusted}
1 (Stepwise)	4213.805	4243.307	0.213
2 (Lasso)	4238.839	4251.483	0.166

- Final Model: Blood Pressure \sim Gender + Age + BMI + Poverty + SleepTrouble
- Mean Prediction Error: 227.951

Model Interpretations

➤ Final Model :

Estimated systolic blood pressure = $96.824 + 4.625 * \text{Gender} + 0.449 * \text{Age} + 0.276 * \text{BMI} - 1.609 * \text{Poverty} - 3.392 * \text{SleepTrouble}$

Predictors : Poverty, Age, BMI, Sleeping quality, Gender

- Binary variables

Gender: 0 -> Female
 1 -> Male

SleepTrouble : 0 -> No
 1 -> Yes

- Numerical variables

Positively related to blood pressure: Age, BMI

Negatively related to blood pressure: Poverty

Recommendations

- Clinics should not advise hypertensive patients who smoke to quit smoking.
- The best predictors of blood pressure are poverty, age, BMI, sleeping quality and gender
- People with obesity problems or who are older need to check their blood pressure regularly.
- The government and medical institutions should educate the public about blood pressure measurement.

Limitations

- Information in the sample data is limited
- Cannot compare predictive power among predictors in the final model

Future Steps

- Explore more potential predictors that can predict blood pressure.
- We can focus on analyzing the association between a particular variable and blood pressure in our final model.

Thank you !