Empirical Analysis of Cardiovascular Data

Harika Prathipati
Sushant Thapa
Lokesh Mylavarpu
Farhan Ar Rafi

Overview

We have analyzed a Cardiovascular dataset and worked with the following predictors of Hypertension Stages:

- Age
- Weight
- Systolic Blood Pressure
- Diastolic Blood Pressure

Based on our analysis we have found a model that can accurately predict Hypertension stage based on the four predictors.

Project Outcomes

- Exploratory Data Analysis (EDA)
- Hypothesis testing
- Correlation
- Linear Regression
- Multiple Linear Regression

Predictor Variables

Age

Age of subjects in years.

Std. Dev - 6.63

Max - 64 yrs

Min - 39 yrs

Mean - 53.49 yrs

Weight

Body weight of subjects in kilograms

Std. Dev - 14.6

Max - 200 kg

Min - 32 kg

Mean - 75.3 kg

Systolic BP

Blood pressure during contraction of the ventricles in

Std. Dev - 14.7

Max - 180 mmHg

Min - 90 mmHg

Mean - 130.26 mmHg

Diastolic BP

Blood pressure recorded just before the next contraction.

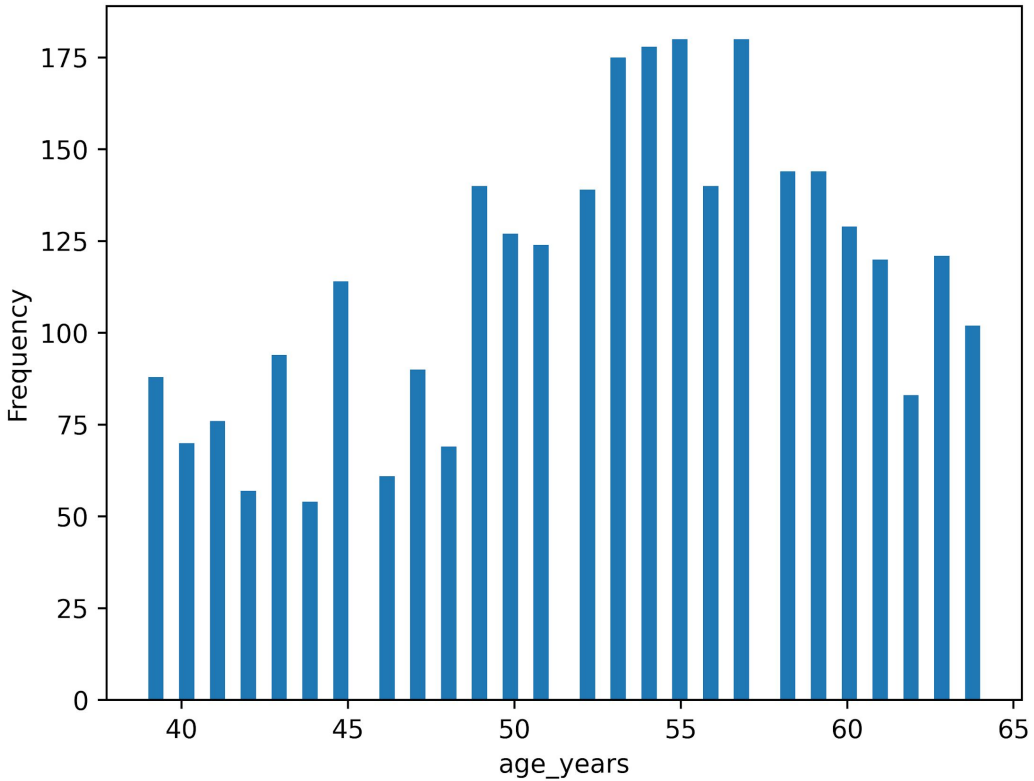
Std. Dev- 7.2

Max - 120 mmHg

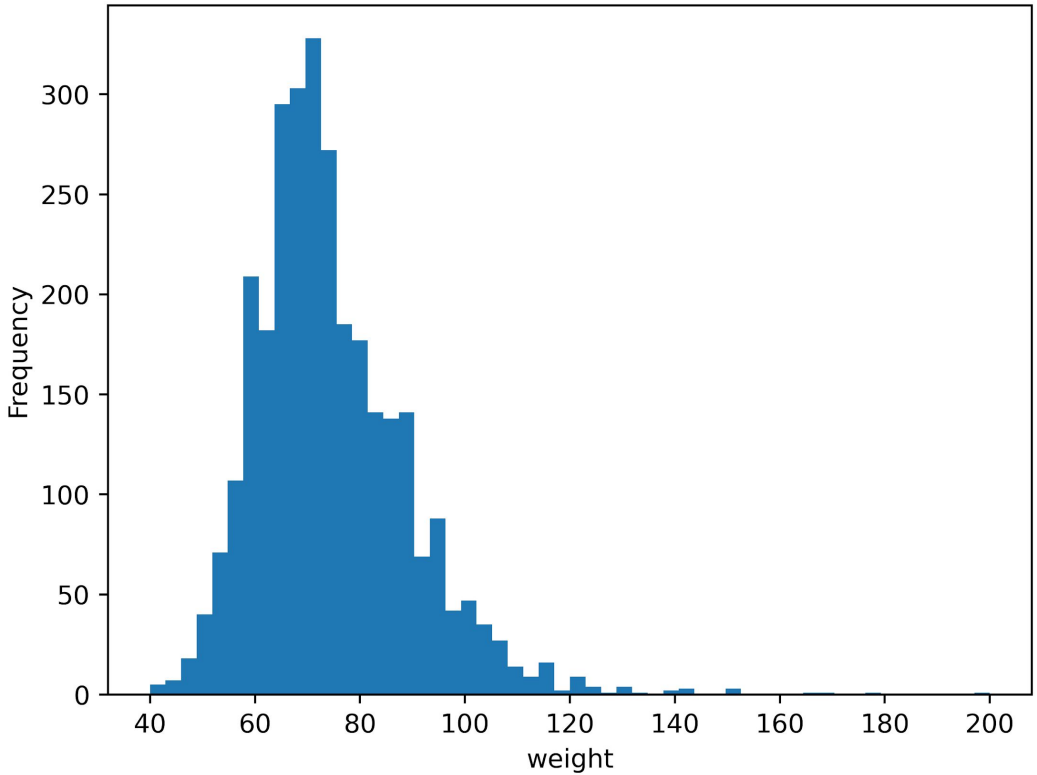
Min -60 mmHg

Mean-84.18 mmHg

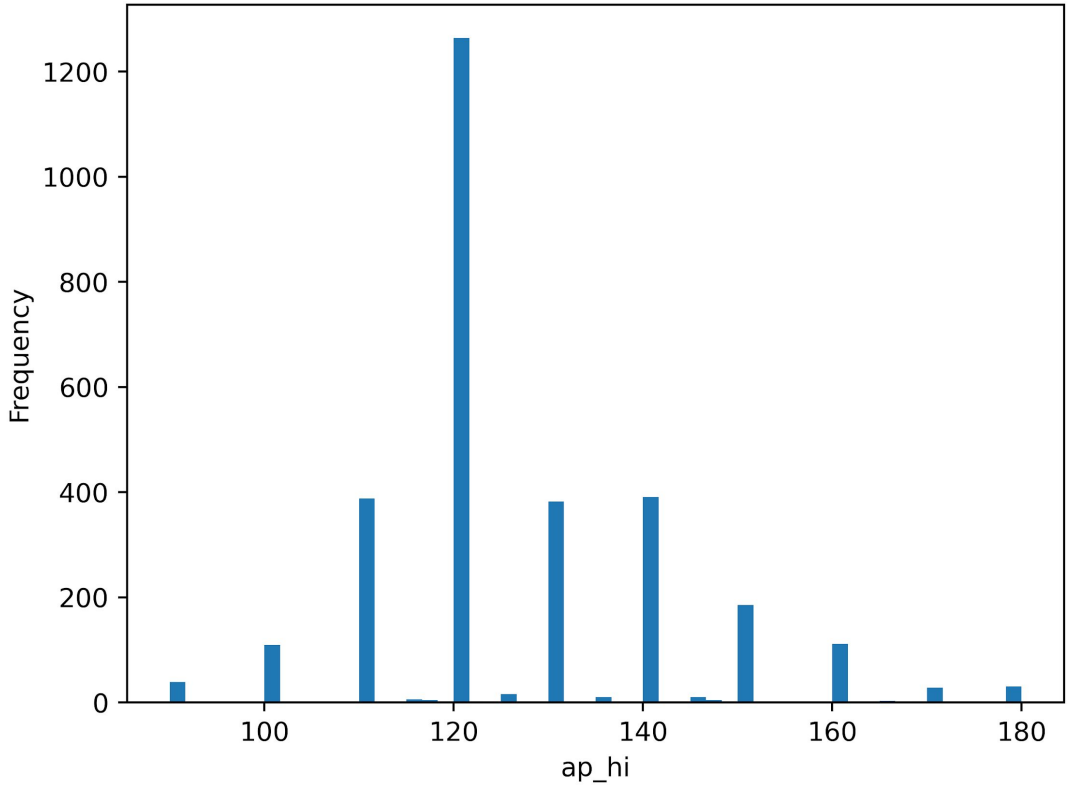
Histogram of Age



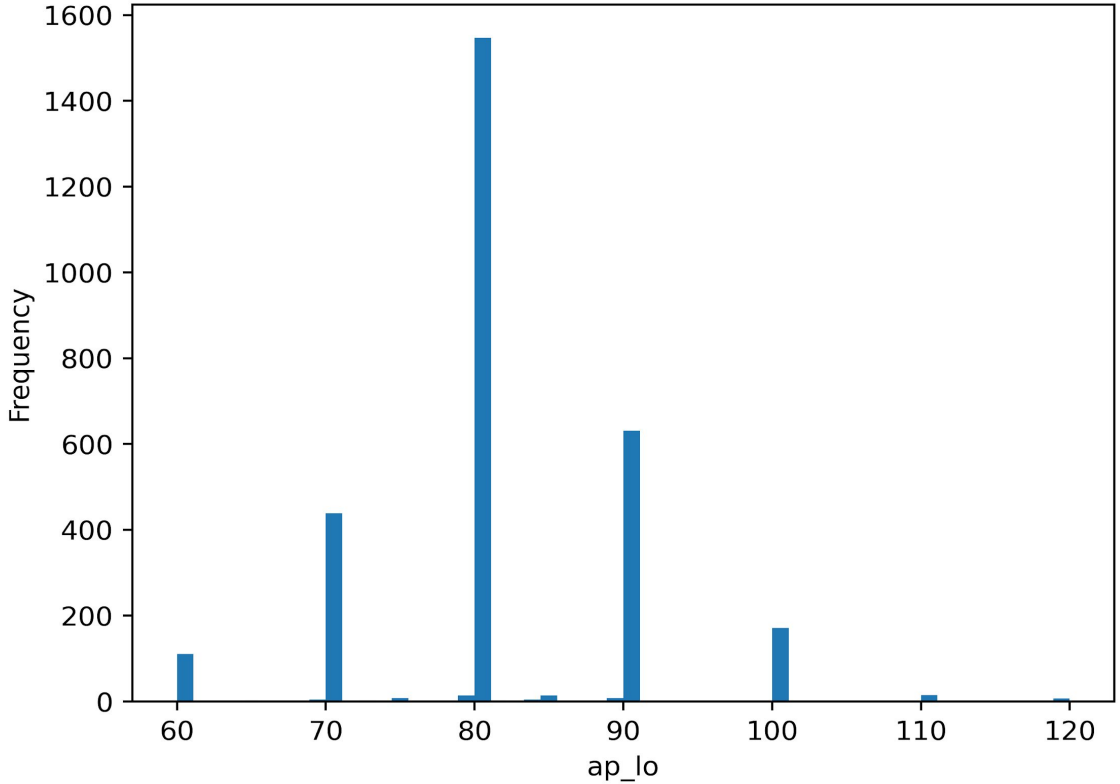
Histogram of Weight



Histogram of Systolic Blood Pressure

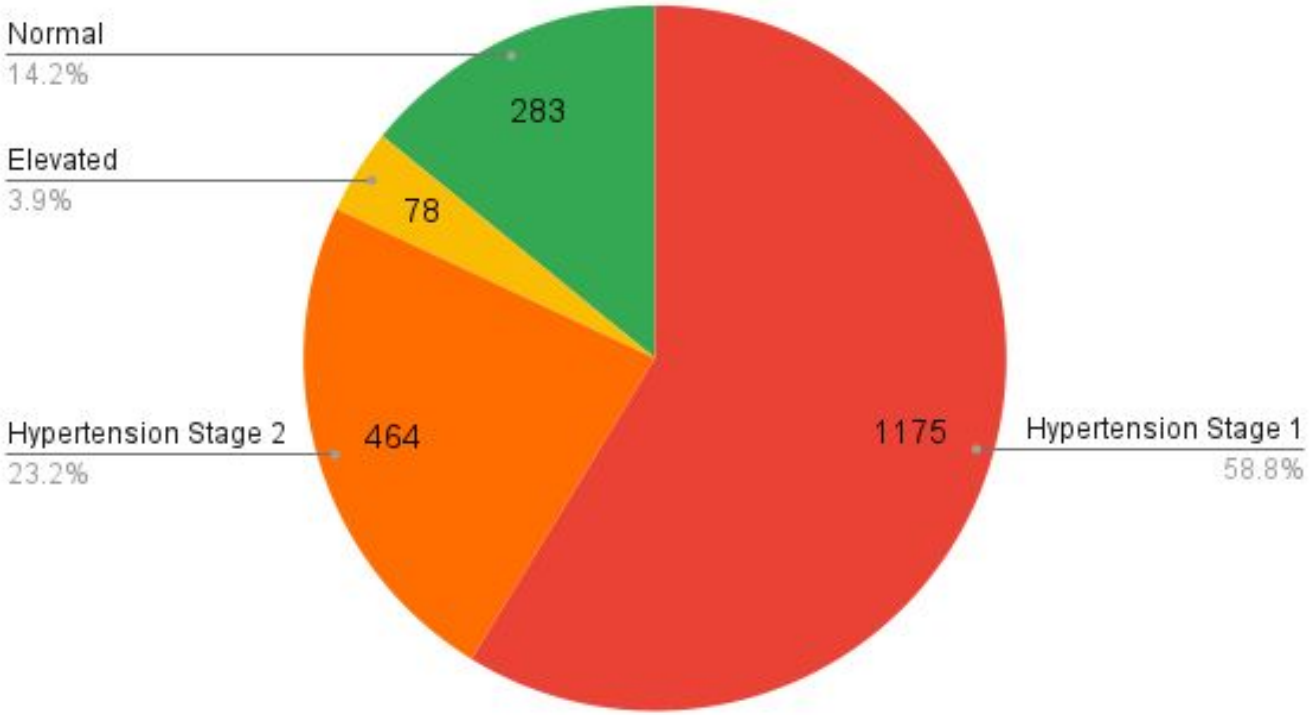


Histogram of Diastolic Blood Pressure



Proportion of Hypertension Stages

Although there are 4 stages in the dataset, we only considered Hypertension Stage 1 and Stage 2 for this experiment.



Hypothesis testing

Approach

1. Collect random sample data. (use python library random)
2. Form a hypothesis (for example, mean age of population is 50 years old)
   a. Guess using sample data
3. Form null and alternative hypothesis.
   Example(for age) :The null hypothesis is H0: $\mu = 50$
   The alternative hypothesis is H1: $\mu \neq 50$
   t_value = 7.034, and critical value is ±1.984
   As t statistic is outside the range given here we reject null hypothesis.
4. Calculate the test statistic according to the case. (using various python libraries like scipy, statistics, and native functions)
5. Calculate the critical values
6. Interpret!

Correlation

Correlation between Predictor Variables

Variable 1	Variable 2	Pearson correlation coefficient
age	weight	0.008
age	ap_hi	0.057
weight	ap_hi	0.108
weight	ap_lo	0.198
ap_hi	**ap_lo**	**0.604**
ap_lo	age	0.13

Verification for high correlation

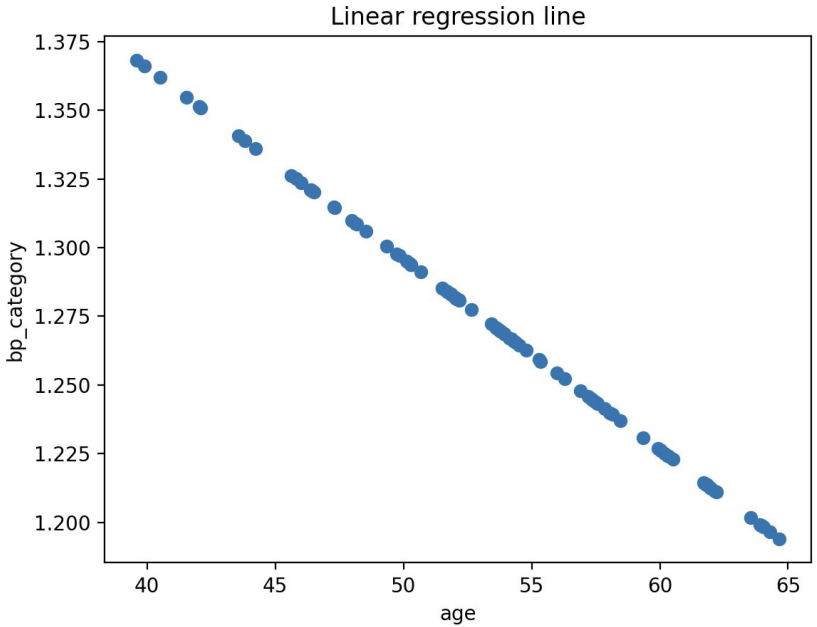
For ap_hi and ap_lo, the confidence interval is between 0.463 and 0.716, which makes sense.

Regression results

Technological Approach

- LinearRegression from sklearn.linear_model
- r2_score from sklearn.metrics for calculating adjusted R^2 score

Age



Result of regression using single predictor variable: **weight**

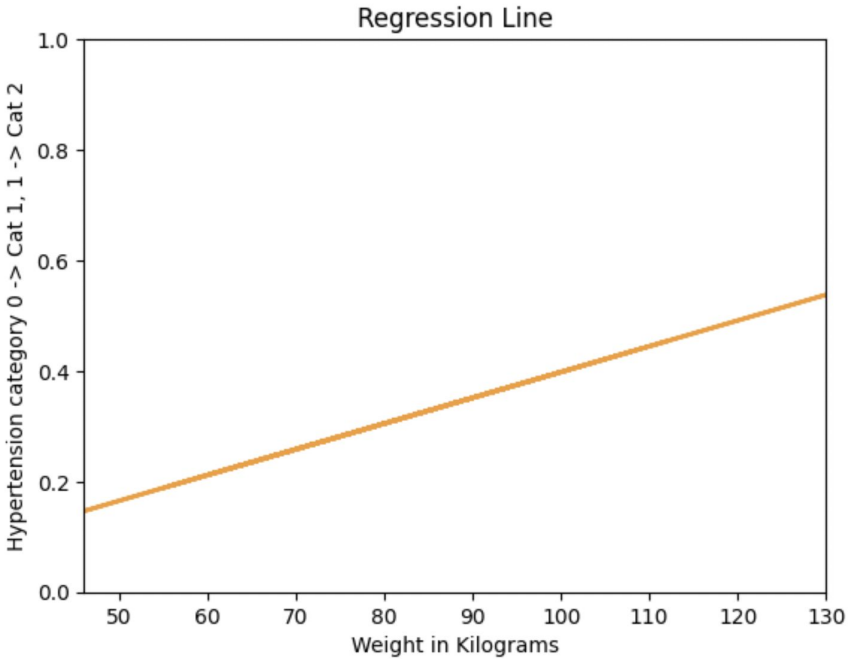
Y-Intercept: $b_0 = -0.0042$

Slope: $b_1 = 1.57$

Determination Coefficient:

$R^2 = 0.00365$

Weight



Result of regression using single predictor variable: **weight**

Y-Intercept: $b_0 = -0.06762726$

Slope: $b_1 = 0.00465721$

Determination Coefficient:

$R^2 = 0.00465721$

Comparison of Predictor Variable combinations

Variables Used	Determination Coefficient R2
Weight **Systolic blood pressure** **Diastolic blood pressure**	0.459
Systolic blood pressure **Diastolic blood pressure**	0.463
Weight **Systolic blood pressure**	0.325
Weight **Diastolic blood pressure**	0.380
Weight	0.0192

Comparison of Predictor Variable combinations

Variables Used	Determination Coefficient R2
age Systolic blood pressure weight	0.461
Systolic blood pressure age	0.382
Weight age	0.037
age	0.380

Comparison of Predictor Variable combinations

Variables Used	Determination Coefficient R2
age **diastolic blood pressure** **weight**	0.674
systloic blood pressure **diastolic blood pressure** **age**	0.701
systloic blood pressure **diastolic blood pressure** **weight**	0.539

Regression results

3 Variables

Result of regression using 3 predictor variables: systolic bp, diastolic bp and age

Determination Coefficient:

R2 = 0.701

2 variables

Result of regression using 2 predictor variables: systolic bp, diastolic bp

Determination Coefficient:

R2 = 0.463

Interpretation

The main determinants of Hypertension:

- Age
- Systolic Blood Pressure
- Diastolic Blood Pressure

Using common sense, the result of data makes sense

Any Questions?