

Prediction of dominating parameters of heart attack in different age groups.

In the present era, heart attack is becoming a life threat especially in young Indians. From published data (www.kaggle.com), I have tried to find out some of the dominating factors of heart attack and finally concluded the probability of those factors.

Google Spreadsheet link of Heart Attack Prediction:

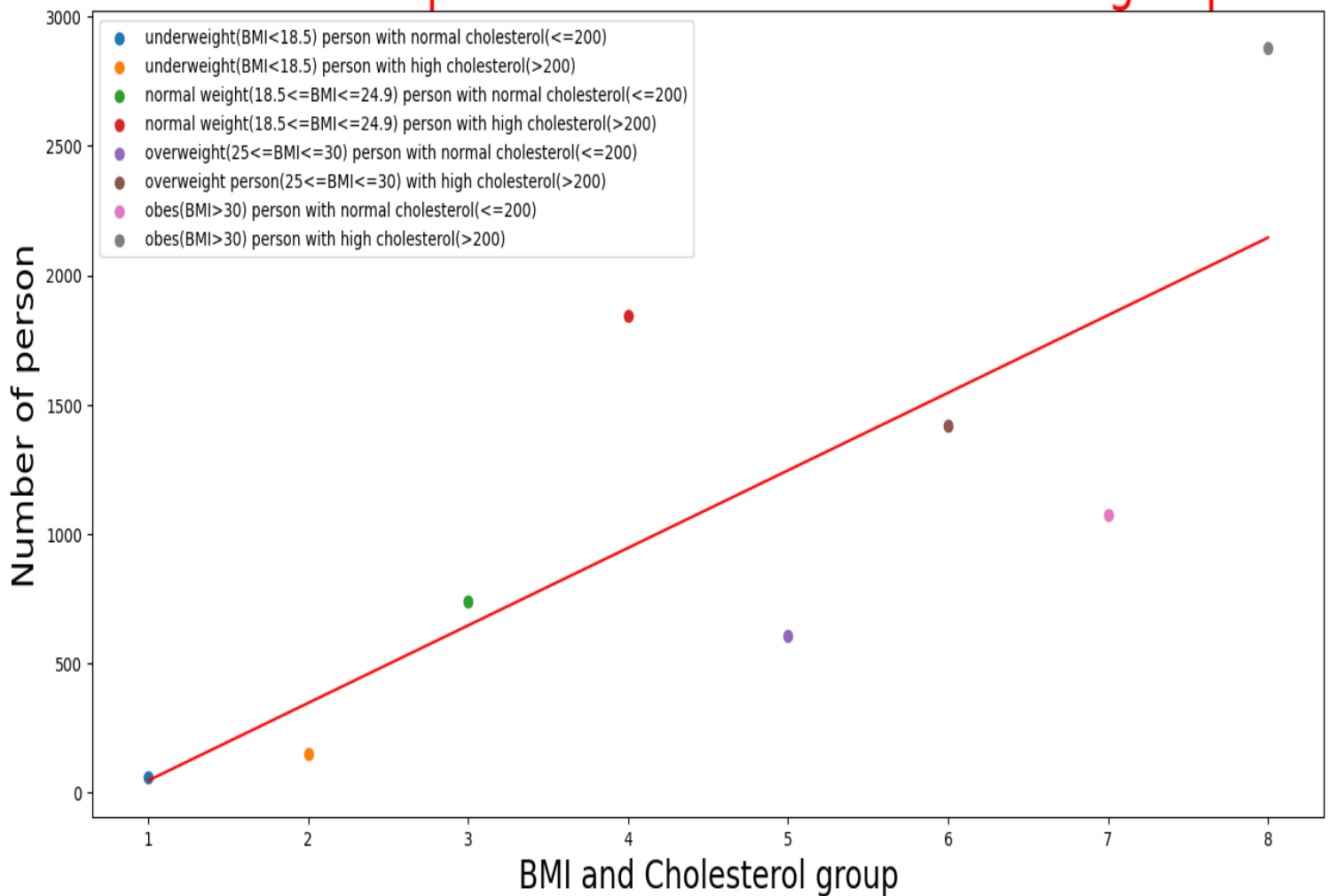
<https://docs.google.com/spreadsheets/d/105y8t8F96KQhXz8vILWgRCTb4IDsy1mjvZ0HYL9EeO8/edit?usp=sharing>

Below are the major parameters related to heart attack

- **Number of people vs. BMI (Body Mass Index) and Cholesterol range**

I have categorised all **8763 people** in **8 sections** with respect to their BMI and cholesterol level such as underweight (BMI $<18.5 \text{ kg/m}^2$) person with normal ($\leq 200 \text{ mg/dL}$) level of cholesterol, underweight (BMI $<18.5 \text{ kg/m}^2$) person with high ($>200 \text{ mg/dL}$) level of cholesterol, normal weight (BMI ~ 18.5 to 24.9 kg/m^2) person with normal ($\leq 200 \text{ mg/dL}$) level of cholesterol, normal weight (BMI ~ 18.5 to 24.9 kg/m^2) person with high ($>200 \text{ mg/dL}$) level of cholesterol, overweight (BMI ~ 24.9 to 29.9 kg/m^2) person with normal ($\leq 200 \text{ mg/dL}$) level of cholesterol, overweight (BMI ~ 24.9 to 29.9 kg/m^2) person with high ($>200 \text{ mg/dL}$) level of cholesterol, obes (BMI $>30 \text{ kg/m}^2$) person with normal ($\leq 200 \text{ mg/dL}$) level of cholesterol and obes (BMI $>30 \text{ kg/m}^2$) person with high ($>200 \text{ mg/dL}$) level of cholesterol. Below is the scatter plot of this data generated using PYTHON language. X axis of this plot represents different classes of BMI and cholesterol level groups (which I have denoted with numbers from 1 to 8) and the Y axis represents the total number of people falling in that certain group.

Number of person vs. BMI and Cholesterol group



Insight from the scatter plot: In the plot, we can see that obese people with high level of cholesterol is maximum in number almost **3000** out of **8763**, underweight person with normal cholesterol level is lowest in number almost **80** out of **8763**. There is a clear trend line with increasing BMI and cholesterol. From the previous calculation almost **36%** of people in the data set have future heart attack risk and almost **50%** of total people have suffered heart disease during their lifetime. This calculation and the trend line in the scatter plot clearly represents that people falling in obese category with high levels of cholesterol have the highest possibility of heart attack in future or they did have some sort of heart disease in their past life. Hence, cholesterol level and BMI of a person is a major parameter for determining heart attack risk in future.

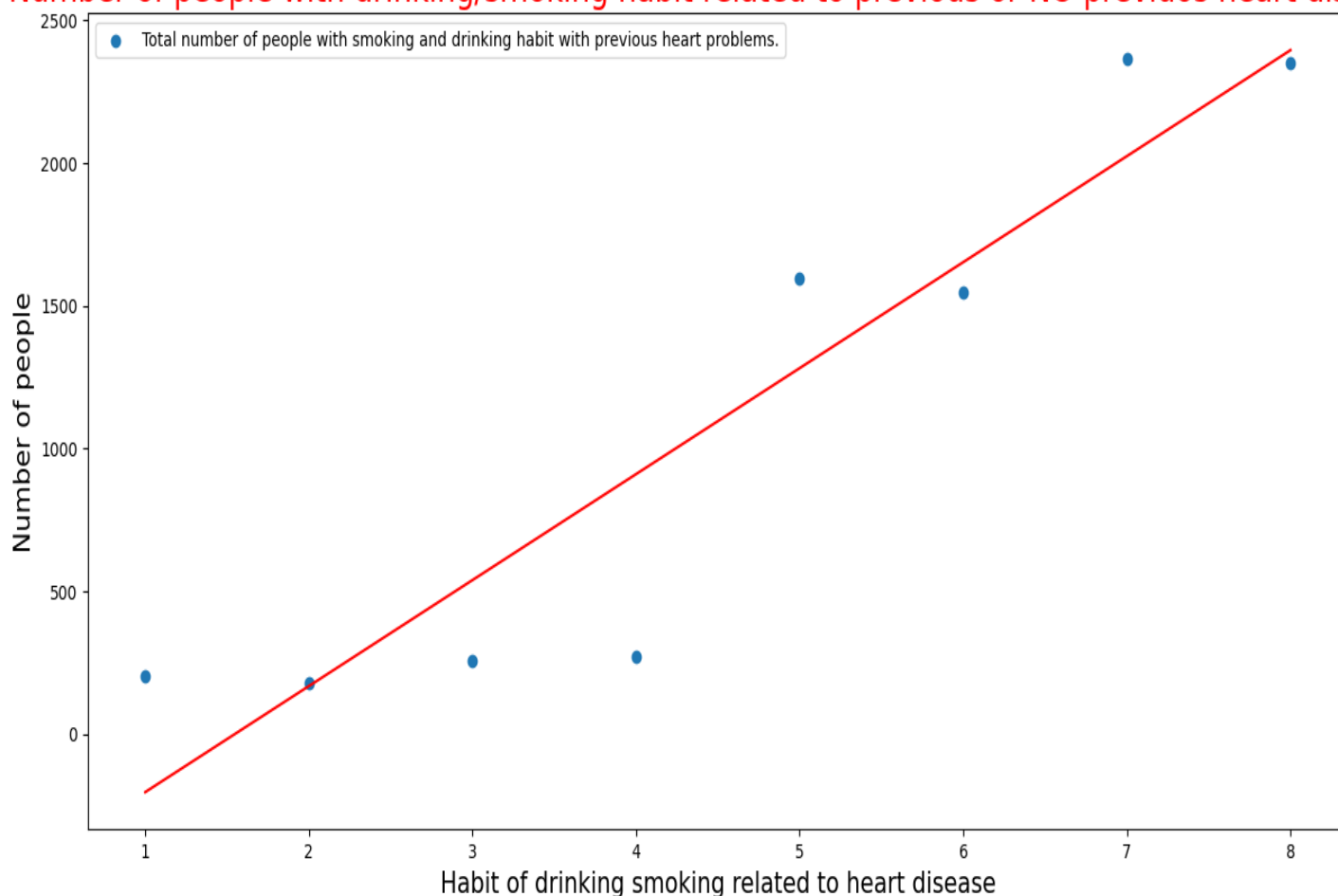
● Number of people vs drinking smoking habit

In this section, I have categorised all **8763 people** in **8 categories** such as people who don't smoke/drink and don't have any previous record of heart disease (2.3%), people who don't smoke/drink but have previous heart disease record (2%), people who don't smoke but drinks and don't have any previous heart disease record

(2.9%), people who don't smoke but drink and have previous heart disease record (3.1%), people who smoke but don't drink and have no previous heart issue record (18%), people who smoke but don't drink and have previous heart issue record (17.6%), people who smoke and drink with no previous heart disease record (27%) and lastly people who smoke and drink with previous heart disease record (27%).

Below is the scatter plot of this data where X axis of the plot represents the 8 classes of people (**denoted by 1 to 8**) discussed so far and Y axis denotes the number of people in a certain class.

Number of people with drinking/smoking habit related to previous or NO previous heart disease.

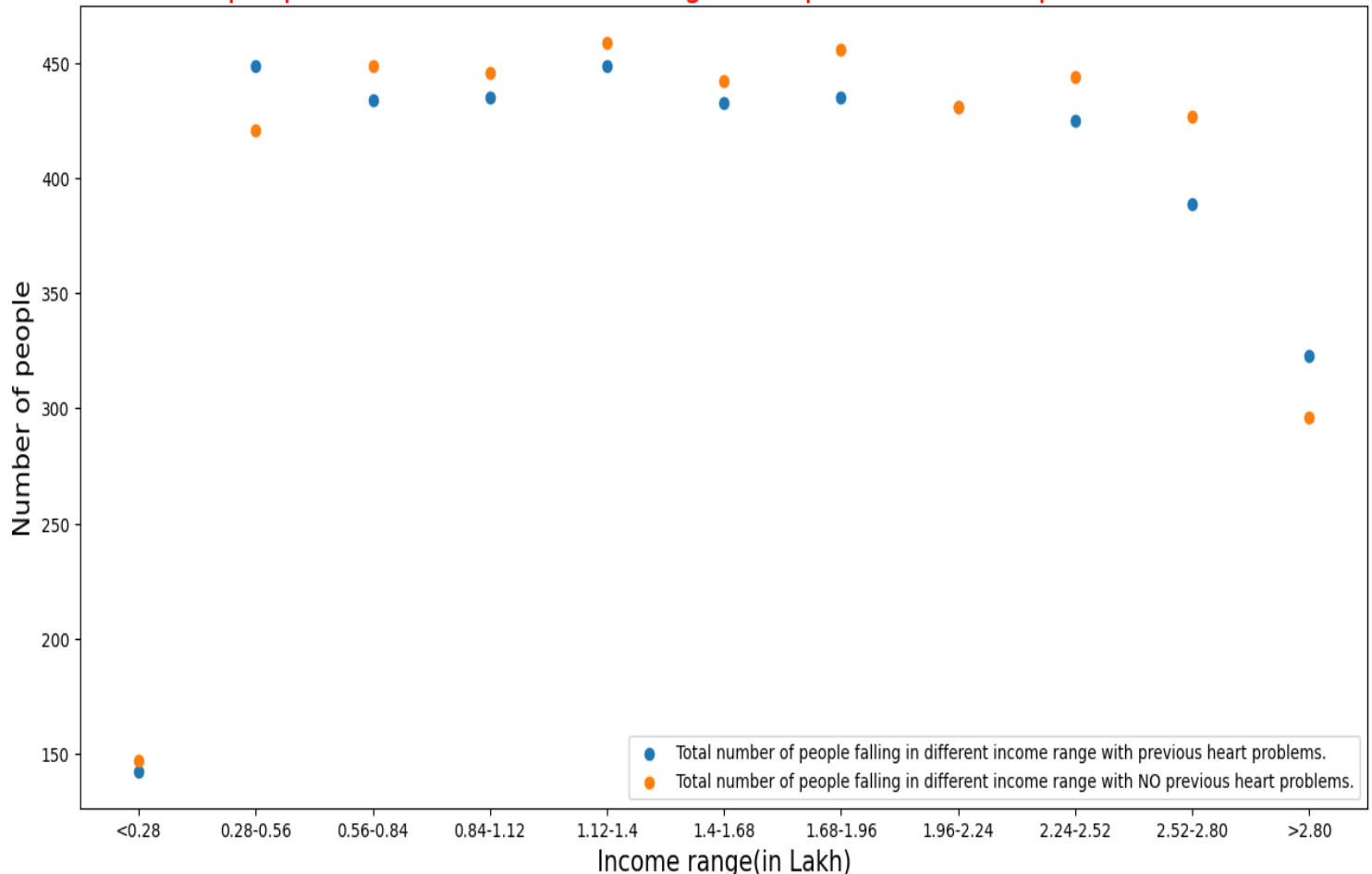


Insight from the scatter plot: From the graph it's obvious that people with smoking and drinking habits and previous heart issues is the highest in number (**27%** of total). On the other hand, only **2%** of total people who don't smoke and drink but have a previous heart issue, that might be a genetic or some other cause. It's also clear that no smoking but only alcohol consumption caused **3.1%** of people to suffer a previous heart disease whereas only smoking but no alcohol caused **17.6%** people to suffer heart disease. Hence, smoking is a major parameter compared to alcohol for future heart risk. Overall, smoking and drinking can be treated as a dominating factor for future heart attack risk.

- **Number of people vs Income range:**

The final plot shows the relation between heart attack risk in people of different income ranges (denoted by 1 to 8 like income <0.28 means $X=1$, income $0.28-0.56$ means $X=2$ and so on $X=8$ means income >2.8 lakh). X axis of the plot corresponds to income range (in **INR lakh**) and Y axis corresponds to the number of people with previous or no previous heart disease records.

Number of people of different income range with previous or NO previous heart disease.



Insight from the scatter plot: It's very clear from the scatter plot that people who have a **very high income (>2.5 lakh)** are **more likely to have a previous heart disease compared to the people with lesser income**. This is quite obvious. People with high income range basically lead a sedentary lifestyle with less physical activity than the people with lesser income. **Data shows almost 50% of people with income >2 lakh, have previous heart disease records.** Hence, income range can be a desired parameter in determining future heart problems.

Calculation of covariance, mean, standard deviation of the above discussed parameters and finding bounds of probability.

The entire calculation to find variance, mean, covariance of different parameters, I have used PYTHON language. Below all the values have been provided.

- **Number of people vs. BMI (Body Mass Index) and Cholesterol range.**

X=BMI and Cholesterol range

Y=Number of people falling in a certain group

Cov(X,Y)=1575.31

Correlation coefficient (ρ) between X and Y=0.78100

Hence, X and Y are strongly correlated.

Standard deviation of X (σ_X)=2.3, mean of X (μ_X)=4.5

Standard deviation of Y (σ_Y)=880, mean of Y (μ_Y)=1095.38

Calculation of bounds:

Let's find out the probability of BMI/Cholesterol range fall between 5 to 8

According to Chebyshev's inequality,

$$P(|X - \mu_X| \geq k\sigma_X) \leq 1/k^2$$

$$5 \leq X \leq 8$$

$$\text{or, } \mu_X - k\sigma_X \leq X \leq \mu_X + k\sigma_X$$

$$\text{Hence, } \mu_X - k\sigma_X = 5 \text{ and } \mu_X + k\sigma_X = 8$$

$$\text{Or, } k = 32/23$$

$$\text{Or, } P(5 \leq X \leq 8) \geq 0.57$$

Which means, at least **57%** cases show overweight or obese category with normal or high cholesterol measurement. **68%** of the total population fall in this category. This indicates a heavy chance of future heart attack among the enlisted persons.

- **Number of people vs. drinking smoking habit.**

X=drinking smoking habit with previous or no previous heart issue

Y=Number of people

Cov(X,Y)=1949.06

Correlation coefficient (ρ) between X and Y=0.94

Hence, X and Y are strongly correlated.

Standard deviation of X (σ_X)=2.29, mean of X (μ_X)=4.5

Standard deviation of Y (σ_Y)=913.35, mean of Y (μ_Y)=1095.375

Calculation of bounds:

Using the same method as before, we have $P(2 \leq X \leq 7) \geq 0.16$

At least **16%** cases ensure that people doing smoking or drinking have a high chance of having future heart issues.

- **Number of people vs. income range.**

$X1$ =Income range

$Y1$ =Number of people with previous heart disease

$X2$ =Income range

$Y2$ =Number of people with NO previous heart disease

$\text{Cov}(X1, Y1)=56$

$\text{Cov}(X2, Y2)=65.54$

Standard deviation of $X1$ (σ_{X1})=3.16 , mean of $X1$ (μ_{X1})=6

Standard deviation of $Y1$ (σ_{Y1})=87.1, mean of $Y1$ (μ_{Y1})=365

Standard deviation of $X2$ (σ_{X2})=3.16, mean of $X2$ (μ_{X2})=6

Standard deviation of $Y2$ (σ_{Y2})=91.33, mean of $Y2$ (μ_{Y2})=401.63

Calculation of bounds:

Using the same method as before, we have $P(8 \leq X \leq 11) \geq 0.60$

At least **60%** cases ensure that **49%** people with income range ≥ 2 lakh have a previous record of heart issue.

SUMMARY:

- We have looked into some dominating factors of Heart attack risk like smoking, drinking habit, BMI and cholesterol level, Income range of people and their lifestyle
- There are many more parameters related to this disease like sedentary lifestyle, Blood pressure, Physical activity, family history, diabetes as well as country of living.
- All these datas should be studied well and research should be performed to develop better treatment for this '**SILENT KILLER**'.