

Data analysis of Chronic disease and health Promotion data

Smoking is the main risk factor for many diseases; In this project, we would like to study the smoking effect (Asthma, high blood pressure, Stroke) , Data used provided by the Centers for Disease Control and Prevention (CDC), Division of Population Health, Epidemiology and Surveillance Branch. With this model we used the smoking data to predict the occurrence of other diseases such as Hypertension, Asthma, and Stroke. after models compared and analyzed, we found that stroke has a higher prevalence rate among smokers compared to the other diseases .

Source of the data:

<https://chronicdata.cdc.gov/500-Cities/500-Cities-City-level-Data-GIS-Friendly-Format-201/dxpw-cm5u>

Upload a csv file
used for the
analysis

Choose CSV File

Browse...

No file selected

☒ Header

Move the slider

Choose Smoking
prevalance(x)

8.510 30.9
8.5 10.9 13.3 15.7 18.1 20.5 22.9 25.3 27.7 30.1 30.9

Smoking rate:

10

Move the slider

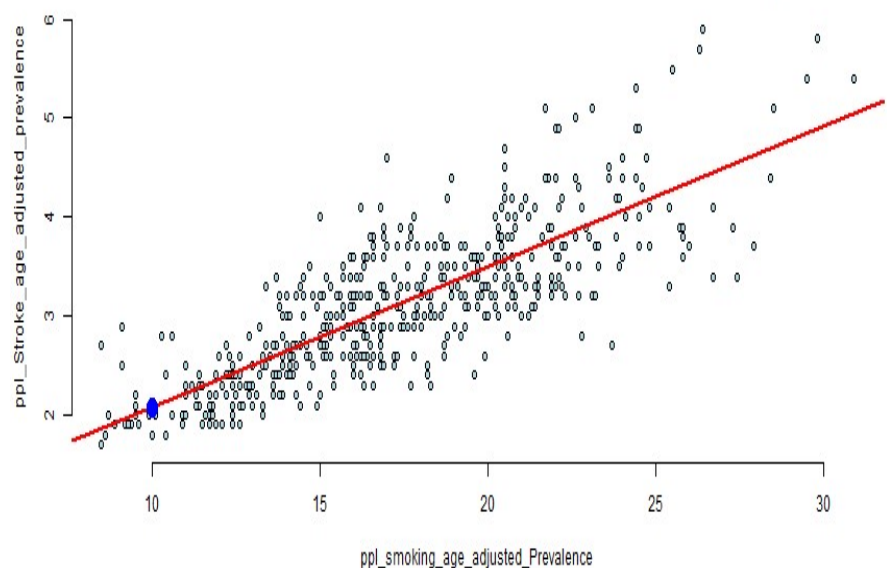
Choose population(x)

43 8,154,190
42 815,146 70,873,261,704,892,536,523,362,154,190

population:

43000

Prediction summary for the residual plot of age adjusted prevalence of smoking among people and age adjusted prevalence of Stroke among people



Prediction1

2.075873

Model Interpretation:

we found this as a good fit because we got adjusted R^2 is greater than others two and close to 1. Values of Residual standard error: 0.4181 on 497, degrees of freedom ,Multiple R-squared: 0.6987, adjusted R-squared: 0.6975, F-statistic: 576.4 on 2 and 497 DF, p-value: $< 2.2e-16$, R^2 is 0.6987. It means data variability in this data is good that is 69.87%

Coefficients Analysis:

Estimate Std.Error t value Pr(>|t|)

(Intercept) -0.232705 0.123020 -1.892 0.0591 .

x 0.112247 0.005436 20.650 $< 2e-16$ ***

ppl_BPHIGH_age_adjusted_prevalence 0.046371 0.004982
9.307 $< 2e-16$ ***

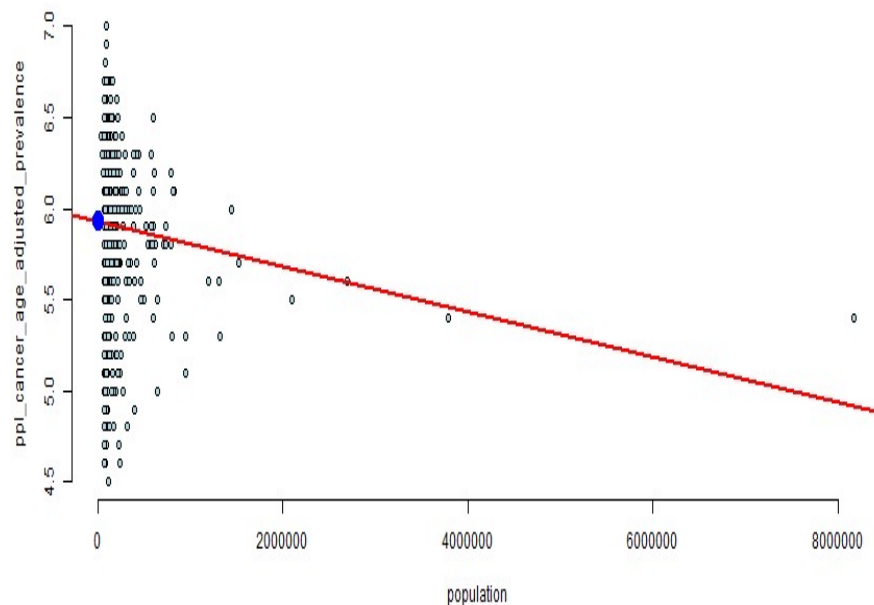
The intercept is the grand mean of stroke rate , calculated for

all data regardless of smoking rate=0, here it is -0.232705

The data here shows that the coefficients for x (smoking rate) is 0.112247. The coefficient indicates that for every additional prevalence of smoke, you can expect stroke prevalence increase by an average of 0.112247 keeping ppl_BPHIGH_age_adjusted_prevalence as constant.

The p-value for x is less than $<2e-16$ here. A very small value means that x is probably an excellent addition to this model.

Prediction summary for the residual plot of population and age adjusted prevalence of cancer among people



prediction2

5.931264

Question: what is the prevalence rate of cancer disease among population?

The data here shows that the coefficients for x (population) is -0.00000012459. The coefficient indicates that for every additional population, you can expect cancer prevalence decreases by an average of -0.00000012459, as it is negative value.