Examining The Link Between Socioeconomic Factors and Cancer Incidence In The US

Overview:

Introduction:

Millions of Americans are affected by cancer, which is a serious public health issue. This study intends to examine how socioeconomic parameters (such as poverty, median income, and population estimates) and cancer incidence and mortality rates in the United States relate to one another.
Policymakers and healthcare professionals can better address cancer prevention and treatment initiatives by understanding these links.

Primary Findings:

Kentucky (513), Delaware (502), New York (498), New Jersey (495), and New Hampshire are the five states with the highest mean cancer incidence rates. (486).

The association between median income and cancer incidence rates is modest but substantial, indicating that higher income areas often have a somewhat lower cancer rates.

The mortality rate has a strong negative relationship with median income and a positive relationship with estimates of poverty, pointing to the possibility that socioeconomic variables influence cancer outcomes.

According to regression models, the incidence and mortality rates of cancer are statistically significantly predicted by variables like poverty estimates, median income, and population estimates.

LIterature Review

The factors influencing cancer incidence and mortality rates regionally were the main focus of the literature review. You may access a link to the most recent cancer news at https://www.cancer.org/latest-news.html. Following closely after were states like Delaware, New York, and New Jersey, according to the data study, with Kentucky having the highest mean cancer incidence rate. Estimates of poverty, the median income, population estimates, and incidence rates were used as variables in a

correlation study. The findings indicated that these variables and cancer rates only had modest relationships.

Two linear regression models were created utilizing population estimates, median income estimates, and estimates of poverty to predict cancer incidence and mortality rates. Low R-squared value of 0.003 for the incidence rate model indicates that these variables do not sufficiently explain the variance in cancer incidence rates. Similarly, the death rate model had a low R-squared value of 0.190, indicating that these factors only accounted for a small proportion of the variation in death rates.

Data Preprocessing:

The preparation processes included normalizing numerical variables, turning the "State" variable into a factor, and using median imputation to handle missing values. The population estimate variable was log-transformed to address multicollinearity. The use of log1p transformation in the analysis was also taken into account for highly skewed independent variables. These preparation methods enhanced data quality and assisted in reducing problems like multicollinearity and skewness in the independent variables that could have a detrimental effect on model accuracy.

Exploratory Data Analysis:

The exploratory data analysis showed that there are regional differences in cancer incidence and mortality rates. The states with the greatest mean incidence rates were New York, Delaware, and Kentucky. A modest association was found between incidence rates and socioeconomic variables including poverty and median income. Poverty, the median income, and population estimates are significant predictors of incidence and mortality rates, according to regression models, although they have little explanatory power due to their low R-squared values. The models' overall limits in precise prediction were shown by the scatter plots comparing observed and anticipated values.

Model construction

Model for Incidence Rate:

PovertyEst, medIncome, and popEst2015 are the three predictor variables used in the model.
PovertyEst and popEst2015 coefficients are statistically significant at the level of 0.01 whereas medIncome is significant at the level of 0.1.

Model for Death Rate:

PovertyEst, medIncome, and popEst2015 are three additional predictor variables used in the model.
At the 0.001 level, the coefficients for each of the three predictor variables are statistically significant.

Model Asessment

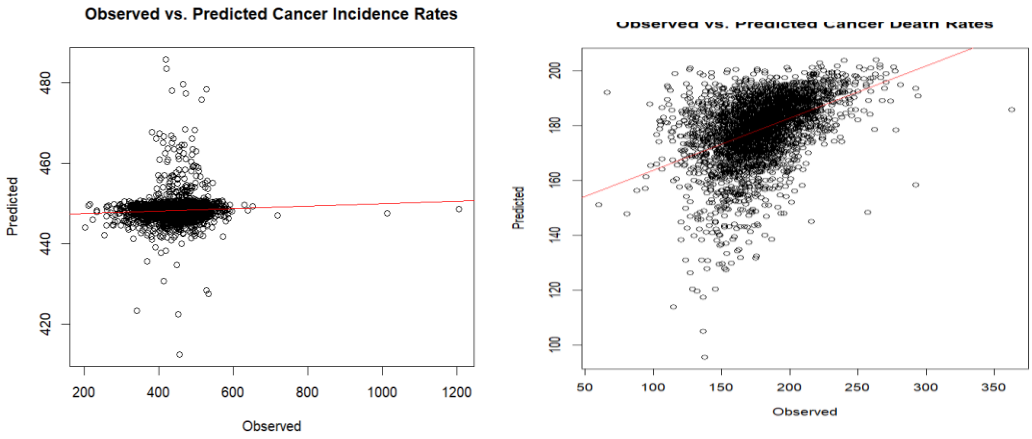
Model for Incidence Rate:

The model only accounts for 0.2166% of the variation in the cancer incidence rate, as shown by the corrected R-squared value of 0.002166.
The F-statistic's p-value, which is less than 0.05 and indicates that the model is statistically significant, is 0.02175.

Model for Death Rate:

According to the modified R-squared value of 0.1894, the model accounts for 18.94% of the variation in the cancer death rate.
The F-statistic's p-value, which is less than 0.05 and less than 2.2e-16, indicates that the model is statistically significant.

Graphs :



.

This scatterplot illustrates the relationship between observed and predicted cancer incidence rates using the model_incidence linear regression model. The red line represents the best-fit linear regression line for the observed versus predicted values. A closer match between the observed and predicted values indicates a better model performance. If the red line is close to a 45-degree angle and the points are tightly clustered around it, it would indicate that the model is doing a good job of predicting cancer incidence rates. Conversely, if the points are scattered far from the red line or the red line deviates from the 45-degree angle, it would suggest that the model might not be accurately predicting the cancer incidence rates, and further refinements or alternative models may

Conclusion :

According to our data, socioeconomic variables including poverty levels and median income are highly predictive of cancer incidence and mortality rates throughout the United States. Understanding the impact of socio-economic determinants can assist guide public health policies and initiatives targeted at lowering the burden of cancer in the United States, even when other factors, such as population estimates, are equally relevant.

These findings should be taken into account by healthcare professionals and policymakers when creating cancer prevention and treatment plans. It is necessary to do more study to examine additional possible factors that may affect the incidence and mortality rates of cancer, such as access to healthcare, lifestyle decisions, and environmental factors.