**Life in America’s Small Cities**

Brashon Ford

National University

**Life in America’s Small Cities**

The data that we have collected tells the story of life in America’s small cities; we are able to know more about their population density, hospital availability, and even income. First let’s look at the data as a whole and find out some exploratory data. The first variable, which is also our independent value, is the death rate per 1000 residents. I found for our analysis that: the mean is 9.30, median is 9.4, standard deviation is 1.66, the maximum is 12.8, and our minimum is 3.6. Our second variable in the dataset is doctor availability per 100,000 residents with the data showing that the mean is 116.0, median is 114, standard deviation is 37.8, minimum is 60, and maximum of doctors is 238. Just by looking at our values we can see that per 100000 residents, there are only 238 doctors available which isn’t a lot. Now tying into death rates and doctor availability, we can look into the hospital availability. The data shows that: mean is 589.8, median is 525, standard deviation is 332.6, minimum is 190, and the maximum is 1792. We can conclude that there is a maximum of 1792 hospitals, with 238 doctors per 100,000 residents.

Let’s take our dataset and perform linear regressions to test the association between X1, dependent variable, and the remaining variables. When we do regression analysis, we also investigate the fitted equation which is Yi=b0+b1Xi+ei. Now we can look at what we found, starting off with holding the death rate per 1000 residents, and comparing it to the doctor availability. Our intercept of that comparison is 8.71 and the increment that we add on to the equation is 0.00508. Next we have the death rate and the hospital availability; our intercept is 8.97 and the b1 is 0.0005. X4 is the variable for annual per capita income, adding that to our dependent variable that data shows: the intercept is 11.81 and the b1 is -0.2659. Lastly, X5 is our variable that interprets the population density of people per square mile. Our analysis came out to be: the intercept is 10.38 and the b1 is -0.0097. I have provided a Figure.1 to visualize the X1 and X2 variables which compare the death rate with doctor availability.

**Figure.1**

**A graph with numbers and lines

Description automatically generated**

Now we can review the collinearity and how it affects our model; collinearity is when two or more independent variables are highly correlated. The correlation between X1 and X2 came out to be 0.115, correlation between X1 and X3 is 0.110, correlation between X1 and X4 is -0.171, and the correlation of X1 and X5 is -0.277. We can see that we have two correlations that are negative values, we know instead of having an upward trend it’ll be downward.

Residual analysis can show us the validity of our regression model, it can identify any patterns or outliers in our analysis. First, I conducted a residual analysis on X1 and X2, which is the death rate per resident and doctor availability. Just by observing the first model of the residuals we can see that there is an outlier that is way past 10, and there’s one that is below -4. I also did another residual analysis for X1 and X3, which is the death rate and hospital availability. Look at Figure.2 to visualize the residuals that are shown with death rate and hospital availability.

**Figure. 2**

**A graph with red line and black dots

Description automatically generated**

**Conclusion**

After doing the analysis and finding necessary data, I would say that this town is small and could use a more doctors in the area. We can use this data to conduct research on what the town needs, what needs to be reevaluated, and how much the people in the area make. This would be a great start to renovating the area that the community lives in.

**References**

 Life in America's Small Cities, by G.S. Thomas