Simple Linear Regression Lab Report
Xinyi (Anny) Cui

1 Letter

Dear Sir or Madam

The purpose of this letter is to seek compensation from your facility for damages to Oregon counties and communities. Our increasing mortality is resulting from the release of radioactive waste from open storage areas into the Columbia River.

We live in a community that has suffered severe casualties and a rapidly rising cancer death rate as a result of the poor oversight of the Handford Nuclear Facility. We Portland and Clatsop are classified as high pollution communities due to the exposure index of radioactive waste exceeding 8. To demonstrate the relative impact of rising cancer mortality rates and waste spills from your storage area, we also combined seven other counties in Oregon, Umatilla, Morrow, Gilliam, Sherman, Wasco, HoodRiver, and Columbia for the tally. Based on the data on waste exposure index and the number of mortality provided by these counties in recent years, we constructed a linear regression model. Its coefficient of determination exceeding 0.8 indicated that 85.81% of the deaths could be reasonably explained by the waste exposure value. Also, the p-value from the t-test of 0.0003321, which is much less than 0.05, proves that the number of deaths has an obvious linear upward relationship with the exposure value. We also tried to use this feasible linear regression model to estimate confidence and prediction intervals, and the results showed that the average mortality in our high-risk community would be as high as 206.9379. Even the low-risk community, consisting of counties with a waste exposure index of less than 4, died The number value will also be as high as 138.0558. Please see more details in the attachment about the discussion on R code.

We hope that the administrators at your facility will take this nuclear waste safety problems seriously and provide reasonable compensation to those in our affected communities. We look forward to hearing from you and would like to schedule one in-person meeting with your facility administrators in mid-October at our community to evalute our specific losses and the amount of compensation.

If your agency staff cannot give our community a satisfactory response, we have the right to appeal to the Oregon state government and complain about your agency's dereliction of duty.

Sincerely yours,

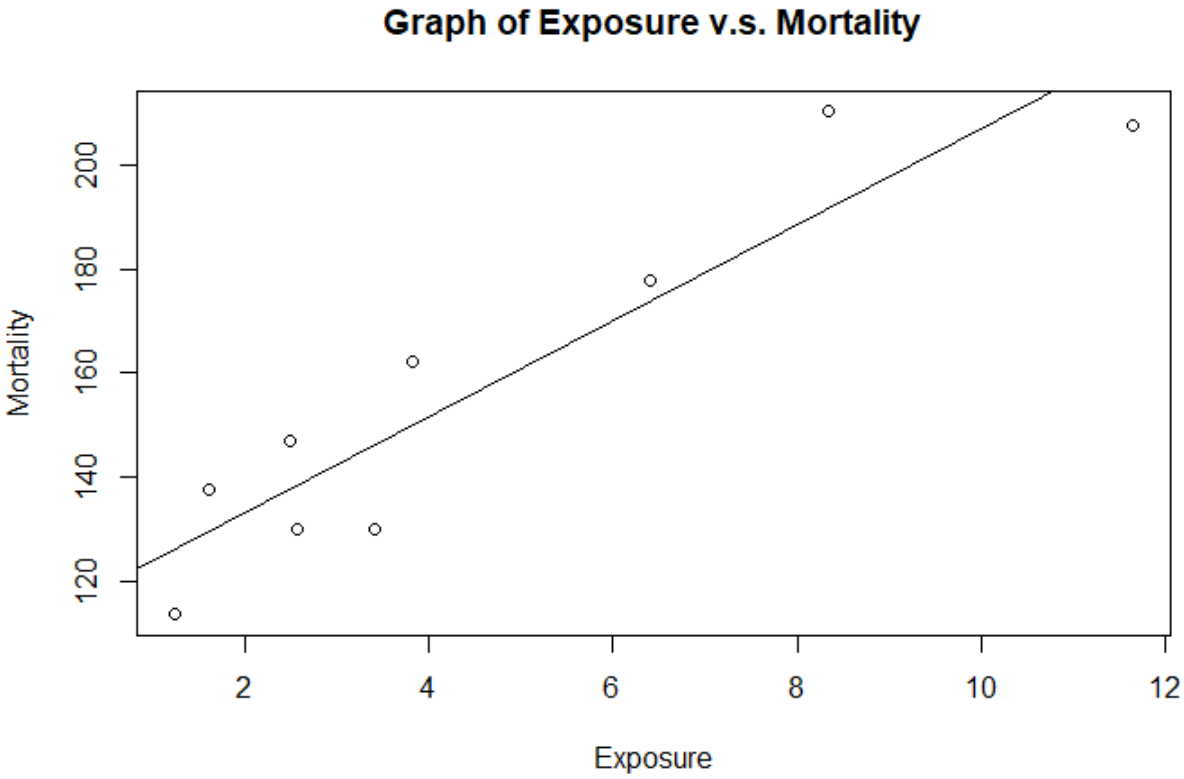
On behalf of Portland

2 Discussion of Model

After making the graph to see the relationship between the amount of exposure and the corresponding morality based on the data given by the website of information "http://www.statsci.org/data/general/hanford.html" and the website of data frame "http://www.statsci.org/data/general/hanford.txt". We can assume there exists a linear relationship between the two variables and then we can construct a linear model shown below:

$$Mortality = 9.231(Exposure) + 114.716$$

The graph with the linear line given by the regression model is shown below:



The coefficient of determination is 0.8581. This means about 85.81% data points of mortality can fall in the regression model line. Over 80% mortality can be explained by the exposure based on the given recorded dataset of counties in Oregon. The p-value of the coefficient is 0.0003321, which means we need to reject H_0 and there exists a clear trend between exposure and mortality..

Then, we can construct a 95% confidence interval which indicates that a range of values with 95% certain contains the population mean. The confidence interval of the slope is $[6.543453, 11.91946]$.
We can assign "low", "medium", and "high" exposure communities base on the amount of their exposure. The counties with less than the value of 4 exposure are in "low" community. The "medium" community contains the counties with the value of 4 to 8. The rest of counties with the value of exposure larger than 8 are "high" community. Then,

Umatilla, Morrow, Gilliam, Sherman, Wasco, and HoodRiver are in the low community; Columbia stays in medium community; High community has Portland and Clatsop. The mean response confidence intervals and prediction intervals for these "low," "medium," and "high" exposure communities are shown below:

```
> rbind(meanLowCI, meanLowPI, meanMediumCI, meanMediumPI, meanHighCI, meanHighPI)
       fit      lwr      upr
1 138.0558 124.9761 151.1356
1 138.0558 102.4390 173.6727
1 173.8893 161.3157 186.4628
1 173.8893 138.4552 209.3233
1 206.9379 185.8007 228.0750
1 206.9379 167.6408 246.2349
```