

Abstract:

This project is designed to find the regional distribution of cancer patients and the impact of the smoking on incident rate and death rate. We're going to do a brief introduction to cancer at first. We choose the three types of data of the United States in the last five years, which are incident rate, death rate and smoking rate. Those are all reliable and good data and examples to do this project. Then, we will use R making the plots and compare those and calculate these data to find the relationship between smoking, cancer incidence and death caused by cancer. Then, we will introduce how we use R code and why we use those to get the result. Finally, we're going to summarize the result and make a conclusion.

Introduction:

Cancer is a serious disease that threatens human life and it's been bothering people for years. Cancer has no limitation, which can start anywhere in the body and it's hard for humans to detect. When humans get cancer, it can affect the normal operation and metabolism of cells and cause eventual death. The United States is a cancer-stricken country and a good example for us to do the research. In this project, we will analyze the information and the relationship between the cancer incident rate and death rate. The data provide us the recent five-year information, we will start to analyze the distribution of the incident and death rate in different counties and we will focus on the effect of age on cancer incidence and mortality of cancer patients since we find age is one of the most significant factors of cancer. Moreover, We're going to compare whether the incidence goes up or down as humans gets older, or vice versa. As countries get richer, more people will develop lifestyle-related cancers than poverty-related cancers. Therefore, it is very necessary to do research on this. Furthermore, we will compare the data and see whether the death rate is decreasing that shows the treatment of cancer is better and better or there is no progress. We hope that these studies will lead to more intuitive data and also help people reduce or prevent cancer.

Data Description:

Death

Data is from CDC's National Program of Cancer Registries Cancer Surveillance System (NPCR-CSS) November 2015 submission, publication of SEER November 2015 submission in United States Cancer Statistics and State Cancer Registry. The data is composed by country, FIPS, Met Objective of 45.5? Age-Adjusted Death Rate, 95% Confidence Interval for Death Rate, Average Deaths per Year, Recent 5-Year Trend and 95% Confidence Interval for Trend.

Incident

Methods:

We first do the ggplot to find the overall relation between data and then the covariance can let us know if there is any relation between the two variables. We try to make the Cp and AIC as small as we can and then take the backward results. We now test the best model from last process and then connect the points from the models. Then we do the test until there is no p-value of any variable is bigger than alpha, we stop the backward process and finally know the results.

Results:

The incident rate decreases a lot these years, but the cancer still cause a lot of people face death. Older person is more likely to get into trouble. It is still believed that cancer can be sealed with medical technology in the future.

Conclusion:

In conclusion, we find incident rate in recent five years are significant decreasing in most counties. However, death rate is mostly stable but also shows a slightly decreased. These result shows cancer prevention in the United States is getting better, and the treatment also make a progress. Moreover, according to the age incidence data, as human get older, the incidence of cancer is higher. So older people are more likely to get cancer than young people, and older people is necessary to prevent cancer. With the development of science and technology and the improvement of medical level, we hope that the mortality rate and incident rate of cancer will continue to decrease as the result obtained from this project in the future.

Appendix:

<http://seer.cancer.gov>

<http://nccd.cdc.gov/usuc/>

http://seer.cancer.gov/stdpopulations/single_age.html

<http://seer.cancer.gov/seerstat/>

<http://www.seer.cancer.gov/stdpopulatin/stdpop.19ages.html>

<http://statecancerprofiles.cancer.gov/datanotavailble.html>

<https://data.world/>

*Data for US does not include Nevada