

## Empirical Model of Heart Disease Risk

This project starts with the use of a 1998 dataset from the Centers for Disease Control that shows the major risks for heart disease and the rate of disease for each state in the U.S (CDCHeartRisk.xls). There are extra labels and a total that you will not want to use in your analysis. You can either import the entire table and then use your program to select the appropriate datasets or you may “clean” the dataset using Excel before you import it.

You should look for the literature on heart disease risks to see what other researchers have out about the risks and the estimates they have made about risk factors. With that information in mind, use the input table to build a multiple regression model where the response variable is heart disease and the possible predictor variables are the risk factors, keeping in mind that these data are summarized at the state level. The risk factors in the table include smoking, obesity, lack of physical activity, and the possible positive impact of eating fruits and vegetables. Test to see which factors are significant at this scale.

### Additional Analyses

1. Once you have a model where you have identified all of the statistically significant predictors, use that statistical equation to build a model that allows the examination of the impacts of policies such as increasing cigarette costs on smoking rates and then on heart disease using the information in the article by Bader et.al. Using their findings, examine the potential reduction in heart disease risks associated with various cigarette cost increases.
2. Assemble a more recent dataset using the links in the references. Rebuild a statistical model to compare it with your original model. Has progress been made in reducing heart disease risk over this time period?

### References

Pearl Bader, David Boisclair, and Roberta Ferrance, Effects of Tobacco Taxation and Pricing on Smoking Behavior in High Risk Populations: A Knowledge Synthesis. *Int. J. Environ. Res. Public Health* 2011, 8(11), 4118-4139; doi:10.3390/ijerph8114118 <http://www.mdpi.com/1660-4601/8/11/4118>.

Kaiser Family Foundation, State Health Facts. <http://kff.org/state-category/health-status/>

Center for Disease Control, National Cardiovascular Disease Surveillance. <https://www.cdc.gov/dhdsp/ncvdss/index.htm>.

Links to more recent data:

<http://kff.org/other/state-indicator/smoking-adults/?currentTimeframe=0&selectedRows=%7B%7D>

Percent who smoke 2015

<http://kff.org/other/state-indicator/number-of-deaths-due-to-diseases-of-the-heart-per-100000-population/?currentTimeframe=0>

heart disease deaths per 100,000

<http://kff.org/state-category/health-status/>

pick out heart disease deaths, smoking, cholesterol, obesity, physical activity