Data Analysis Project 2

Due: Mar 26, 2015 in class

Identification of Risk Factors for Heart Disease Using Wisconsin Longitudinal Study Data

Background

Heart disease is the leading cause of deaths worldwide. Although heart disease usually affects older adults, the antecedents of disease, notably atherosclerosis, begin in early life, making primary prevention efforts necessary from childhood. Therefore, there is increased emphasis on identifying risk factors for heart disease and developing intervention to modify risk factors for young people.

Study

The Wisconsin Longitudinal Study (WLS) is a long-term study of a random sample of 10,317 men and women who graduated from Wisconsin high schools in 1957. The WLS provides an opportunity to study the life course, intergenerational transfers and relationships, family functioning, physical and mental health and well-being, and morbidity and mortality from late adolescence through 2008. WLS data also cover social background, youthful aspirations, schooling, military service, labor market experiences, family characteristics and events, social participation, psychological characteristics and retirement.

See more information at http://www.ssc.wisc.edu/wlsresearch/

Objectives

Identify risk factors for heart attack (a specific heart disease).

Data Descriptions

The data used in this project contain the survey data from graduates and their siblings in 1957, 1993, 2004 and 2011. For each graduate, one sibling (if any) was selected to participate in the study. The variable "idpub" is the ID for the (graduate, sibling) pair. The variable "Rtype" indicates the observation was from the graduate (if Rtype=g) or from the sibling (if Rtype=s). For each variable, the end of the name indicates the year the variable was measured. For example, "HA2011" means the variable was measured in 2011. If there is no year in variable name, it means that the variable was measured in 1957 (for example, variable "sex" and "IQ").

The detailed meaning of each variable can be found in "Variable Descriptions.pdf".

Response: HA2011, HA2004, doc2011, doc2004 (There may be a conflict between HA and doc for some subjects.)

Covariates: All other variables. The covariates are divided into several categories, such as general health_weight, general health_sleep, family's medical condition, etc. Please refer to "Variable Categories.pdf" for details.

Meetings with the Researcher

We will meet twice with the researcher to have the opportunity to ask questions. These meetings will be during class on March 19 Thursday and March 24 Tuesday.

Data Analysis and Report

You are to conduct an analysis of the data for the two aims above and write a report directed to the researcher. The report is due in class on Mar 26 Thursday.

The maximum length for your report is 12 pages; this includes title, abstract, introduction, data, methods, results, and discussion. The report text must be double-spaced and in a font that produces no more than 32 lines of text per page (e.g., 1.5 line spacing, standard width, and various 11pt fonts in Word; or various LaTeX formats using 11pt or 12pt). Integrate key figures and tables (with captions) into the report.

Information critical to major analyses should appear in the main body of the report. A very brief Appendix A for supplementary materials is not necessary but may be included. Include your lists of questions for the researcher as Appendix B. Include your computer code in Appendix C. Appendixes are not counted toward the 12-page limit. Number all the pages on your report including the appendixes.

Your analysis should be thorough, correct, and as straightforward as possible. Your writing should be thorough, accurate, clear, and concise, while containing the appropriate amount of "hard" statistical information. Finding the balance between too much and too little formal statistical information is a key part of report preparation.