**BS730 – Spring 2020**

**Project 2**

**Due Date: Tuesday 28th April by 10am (beginning of class)**

**Details**

This project must be completed individually – no collaboration at all is allowed between students on all aspects of the project. The project is worth 15% of your final grade. We encourage you to work as independently as possible on the project although you may ask the teaching team for assistance with programming issues when absolutely necessary. Your project submission should include completed tables, write-up and any graphs for each part. Your R code needs to be easily readable (neat and tidy), clearly labelled and included as an Appendix. Please submit on Blackboard by the beginning of class on 28th April 2020 by 10am before class begins. Points are assigned for presentation and organization. Projects must be a maximum of 5 single-sided A4 pages (which includes the R Code appendix). Late submissions are penalized 5 points per day unless prior arrangements have been made with your course instructor. No late projects will be accepted after three days past the due date and time.

**Background**

The Jackson Heart Study (JHS), which initiated in 1998, is a single-site, prospective, epidemiologic investigation of cardiovascular disease among African Americans. The JHS is funded by the National Heart, Lung and Blood Institute (NHLBI) and the National Institute on Minority Health and Health Disparities (NIMHD). It is a population-based longitudinal study. Data is available from multiple participant examinations, denoted as visit 1 through visit 3. The JHS recruited 5306 African American residents living in the Jackson, Mississippi, metropolitan area of Hinds, Madison, and Rankin Counties. Participants were enrolled from four recruitment pools: random, 17%; volunteer, 30%; currently enrolled in the Atherosclerosis Risk in Communities Study (ARIC), 31% and secondary family members, 22%. Recruitment was limited to non-institutionalized adult African American men and women, 35-84 years old, except in a nested family cohort where those 21 to 34 years of age were also eligible. The final cohort of participants enrolled during the baseline exam included 6.6% of all African American men and women residents of the Jackson Mississippi Metropolitan Statistical Area aged 35-84 (N=76,426, US Census 2000). Because there is a greater prevalence of cardiovascular disease among African Americans compared to other ethnic groups, the purpose of the Jackson Heart Study is to explore the reasons for this disparity and to uncover new approaches to reduce it. (1)

Life's Simple 7 is defined by the American Heart Association as the 7 risk factors that people can improve through lifestyle changes to help achieve ideal cardiovascular health. (2) These include managing blood pressure, controlling cholesterol, reducing blood sugar, being active, eating better, losing weight and stopping smoking. There are multiple ways of evaluating these 7 risk factors. One is awarding an individual one point for each category for which they have “ideal health” resulting in a score that would range from 0 to 7. While numerous studies have examined Life’s Simple 7 in terms of cardiovascular disease (see (3) for a meta analysis) there is also interest in how 6 (“Simple 6”) of these factors relate to diabetes. (For studies of diabetes blood sugar is not included as it is a clinical measure used in the diagnosis of diabetes.) The relationship between the Simple 6 and incident diabetes has been examined in the Jackson Heart Study (4) longitudinal data. In this project the main focus will be on the association between the Simple 6 and its component measures with the outcomes of diabetes and the continuous measure of fasting plasma glucose among individuals without diabetes at visit 3.

Project 2 will be an extension of the analysis completed in Project 1 in that it also uses JHS data and focuses on Life’s Simple 6. You will again use a subset of the JHS TRANS-Data Package, which consists of a de-identified build of JHS data(5). Although the topics are related, you are provided a new data set for project 2 and you do not need to refer to the analysis you performed for Project 1.

**The aims of Project 2 are:**

1. To examine the association between type of insurance and two outcome variables measured at visit 3: fasting plasma glucose level among individuals without diabetes, and diabetes status.
2. To examine the association between various risk potential risk factors (demographics, lifestyle, clinical, and environmental) and the outcomes
   1. fasting plasma glucose level (among individuals without diabetes)
   2. diabetes
3. To assess the presence of effect modification by sex on the association between the exposures of bmi and total cholesterol and the outcome of fasting plasma glucose in individuals without diabetes.

The dataset is available in the file **jhst\_proj2final.csv**. A description of the variables in the dataset can be found on page 3. Any missing data is denoted by a blank space.

**References**

1. Wyatt SB, Diekelmann N, Henderson F, Andrew ME, Billingsley G, Felder SH et al. A community-driven model of research participation: the Jackson Heart Study Participant Recruitment and Retention Study. Ethn Dis 2003; 13(4):438-455.
2. <https://www.heart.org/en/healthy-living/healthy-lifestyle/my-life-check--lifes-simple-7>
3. *Guo L, Zhang S. Association between ideal cardiovascular health metrics and risk of cardiovascular events or mortality: A meta-analysis of prospective studies. Clin Cardiol. 2017;40(12):1339–1346. doi:10.1002/clc.22836*
4. Effoe VS, Carnethon MR, Echouffo-Tcheugui JB, et al. The American Heart Association Ideal Cardiovascular Health and Incident Type 2 Diabetes Mellitus Among Blacks: The Jackson Heart Study. *J Am Heart Assoc*. 2017;6(6):e005008. Published 2017 Jun 21. doi:10.1161/JAHA.116.005008
5. <https://www.jacksonheartstudy.org/Research/Data-Science>

**Data Dictionary**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Variable Type** | **Description** | **Coding** |
| subjid | Numeric | Participant ID | Integer value between 1 and 2750 |
| ageV3 | Numeric | Participant age at visit 3 | Range 31.9 to 89.8 years |
| sex | Numeric | Participant sex | Female = 0  Male = 1 |
| BMIV3 | Numeric | Body mass index (kg/m2) at visit 3 | Range 18.23 to 59.55 kg/m2 |
| DMmedsV3 | Character | Takes medication for diabetes at visit 3 | Yes  No |
| sbpV3 | Numeric | Systolic blood pressure (mmHg) at visit 3 | 90.00 to 194.00 mmHg |
| dbpV3 | Numeric | Diastolic blood pressure (mmHg) at visit 3 | 50.00 to 108.00 mmHg |
| HTNV3 | Character | Hypertension at visit 3 | Yes  No |
| FPGV3 | Numeric | Fasting plasma glucose (mg/dL) at visit 3 | 69.00 to 302.00 t mg/dL |
| DiabetesV3 | Character | Diabetes at visit 3 | Yes  No |
| totcholV3 | Numeric | Fasting total cholesterol (mg/dL) at visit 3 | 106.00 to 328.00 mg/dL |
| incomev3 | Character | Income category | Affluent  Upper-middle  Lower-middle  Poor  Unknown |
| idealhealthSMKv3 | Numeric | Ideal Health for Smoking | 0= Not ideal Health  1= Ideal Health  For smoking |
| IdealHealthPAV3 | Numeric | Ideal Health for Physical Activity | 0= Not ideal Health  1= Ideal Health  For physical activity |
| idealhealthNutritionv3 | Numeric | Ideal Health for Nutrition | 0= Not ideal Health  1= Ideal Health  Nutrition |
| PrivatepublicInsV3 | Character | Public or Private Insurance | Uninsured  Public Only  Private Only  Public & Private |
| Simple6 | Numeric | Sum of Ideal Health factors over BMI, Blood Pressure, Total Cholesterol, Smoking, Physical Activity, Nutrition | Possible range 0 to 6  Observed range 0 to 5  **Treat as a continuous variable in all analyses** |

**General Instructions:** For all parts of this project, analyses of fasting plasma glucose should exclude individuals who are diabetic. If there are individuals with missing diabetic status they can remain in the fasting plasma glucose analysis. For diabetes status, model the outcome of “Yes” as the outcome of interest. For all binary variables coded as 0/1, including sex, use 0 as the reference group.

**Part 1. Data Analysis for Aim 1 (20 points) .** The objective of Aim 1 is to examine the association between type of insurance and two different outcomes as follows (i) the continuous measure related to diabetes of fasting plasma glucose level and (ii) the binary variable of diabetes status. Complete the following analyses and use your results to fill in Table 1.

1. Perform an appropriate statistical analysis to assess the association between type of insurance and the outcome of fasting plasma glucose at visit 3 excluding individuals with diabetes. Fit a model with just type of insurance and a separate model with type of insurance adjusted for age and sex. As part of your analysis evaluate whether or not post-hoc tests are appropriate.
2. Perform an appropriate statistical analysis to assess the association between type of insurance and the outcome of diabetes status at visit 3. Fit a model with just type of insurance and a separate model with type of insurance adjusted for age and sex.

**Part 2a. Data Analysis for Aim 2: Fasting Plasma Glucose (20 points).** The first objective of Aim 2 is to examine the association between various risk factors and fasting plasma glucose among JHS participants excluding diabetic participants. Complete the following analyses and use your results to fill in **Table 2a** (all spaces marked ???).

1. Perform appropriate analyses to assess the association between each exposure variable shown in Table 2a, and also for fasting plasma glucose at visit 3 excluding diabetic individuals. Each model should be adjusted for age and sex.
2. Repeat the analysis in part a using a single multivariable adjusted model that includes all of the exposure variables in Table 2a except Simple 6. The effect measure for each exposure should be adjusted for all of the other variables listed in Table 2 (except Simple 6).

**Part 2b. Data Analysis for Aim 2: Diabetes (20 points).** A second objective of Aim 2 is to examine the association between various risk factors and diabetes at visit 3 among JHS participants. Complete the following analyses and use your results to fill in **Table 2b** (all spaces marked ???).

1. Perform an appropriate analysis to assess the association between the exposure variables shown in Tables 2b and the outcome of diabetes at visit 3. All models should be adjusted for age and sex.
2. Repeat the analysis in part a using a multivariable adjusted model that includes all of the exposure variables in Table 2b except Simple 6. The effect measure for each exposure should be adjusted for all of the other variables listed in Table 2 (except Simple 6).

**Part 3. Data analysis for Aim 3 (20 points).** The objective of Aim 3 is to assess the presence of interaction between sex and the following exposure variables: total cholesterol, and BMI for the measure of fasting plasma glucose among non-diabetic individuals in the JHS. Complete the following analyses and use your results to fill in **Table 3** (all spaces marked ???).

1. Perform a formal test of interaction between each exposure variable (total cholesterol, and BMI) and sex, adjusting for age, on the outcome of fasting plasma glucose excluding diabetic individuals. Do not include any additional covariates in the model. Specifically, fit the following two models:
   1. Model 1: Outcome: fasting plasma glucose excluding diabetics. Independent variables: age, sex, BMI, total cholesterol, sex and total cholesterol interaction
   2. Model 2: Outcome: fasting plasma glucose excluding diabetics. Independent variables: age, sex, BMI, total cholesterol, sex and BMI interaction

**Part 4. Write-up (20 points).**

1. Write a brief (maximum one-half page) summary of the statistical methods used to create Tables 1 (a&b), 2 (a&b), and 3. Do NOT give specific R procedures (e.g. t.test, etc.) or a full list of all analyses you did for each variable.
2. Write a results/conclusion section that summarizes the main findings from Tables 1-3. Comment on any interesting findings and summarize implications of your results. The entire section should be no more than one-half page long.

**Please include your R code in an Appendix at the end of your write-up. DO NOT INCLUDE R OUTPUT.**

**Table 1a. Association between fasting plasma glucose and type of insurance.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Outcome = Fasting Plasma Glucose among non-diabetic individuals at visit 3**  **Number of Observations included in analysis:** ??? | | | | | |
| **Model 1 – Unadjusted** | | | **Model 2 – Multivariable adjusted\*** | | |
| **N** | **β Estimate and Std error** | **T-test p-value** | **N** | **Adjusted β Estimate and Std error** | **T-test p-value** |
| **Private/Public Insurance** |  |  |  |  |  |  |
| Public & Private | ??? | Referent |  | ??? | Referent |  |
| Private Only | ??? | ??? | ??? | ??? | ??? |  |
| Public Only | ??? | ??? | ??? | ??? | ??? |  |
| Uninsured | ??? | ??? | ??? | ??? | ??? |  |
| Age |  |  |  | ??? | ??? |  |
| Sex |  |  |  | ??? | ??? |  |

\* Adjusted for age, sex and type of insurance

**Table 1b. Association between diabetes status and type of insurance.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Private/Public Insurance** | **Outcome = Diabetes at visit 3**  **Number with Diabetes** ??? **Number without Diabestes** ???  (based on individuals used in the multivariate analyses) | | | |
| **Model 1 – Unadjusted** | | **Model 2 – Multivariable adjusted\*** | |
| **OR (95% CL)** | **Z-test P-value** | **OR (95% CL)** | **Z-test P-value** |
| Public & Private | Referent |  | Referent |  |
| Private Only | ??? | ??? | ??? | ??? |
| Public Only | ??? | ??? | ??? | ??? |
| Uninsured | ??? | ??? | ??? | ??? |
| Age |  |  | ??? | ??? |
| Sex |  |  | ??? | ??? |

\* Adjusted for age, sex and type of insurance

**Table 2a.** **Association between Simple 6 and its individual components with fasting plasma glucose.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Exposure** | **Outcome = Fasting Plasma Glucose among non-diabetic individuals at visit 3**  [Total participants = ???] | | | |
| **Model 1 – Age/sex adjusted** | | **Model 2 – Multivariable adjusted\*** | |
| **β estimate ± SE** | **P-value** | **β estimate ± SE** | **P-value** |
| Age (years) | ??? | ??? | ??? | ??? |
| Sex | ??? | ??? | ??? | ??? |
| Body mass index (kg/m2) | ??? | ??? | ??? | ??? |
| Systolic Blood Pressure (mmHg) | ??? | ??? | ??? | ??? |
| Diastolic Blood Pressure (mmHg) | ??? | ??? | ??? | ??? |
| Total Cholesterol (mg/dL) | ??? | ??? | ??? | ??? |
| Ideal Smoking Health | ??? | ??? | ??? | ??? |
| Ideal Physical Activity Health | ??? | ??? | ??? | ??? |
| Ideal Nutrition Health | ??? | ??? | ??? | ??? |
|  |  |  |  |  |
| Simple 6 | ??? | ??? | NA | NA |

\*The β estimate for each exposure is adjusted for all of the other exposure variables listed in the table (except Simple 6).

**Table 2b.** **Association between Simple 6 and its individual components with diabetes status.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Exposure** | **Outcome = Diabetes at visit 3**  **Number with Diabetes** ??? **Number without Diabetes** ???  (based on individuals used in the multivariate analyses) | | | |
| **Model 1 – Age/sex adjusted** | | **Model 2 – Multivariable adjusted\*** | |
| **OR**  **(95% CI)** | **P-value** | **OR**  **(95% CI)** | **P-value** |
| Age (years) | ??? | ??? | ??? | ??? |
| Sex | ??? | ??? | ??? | ??? |
| Body mass index (kg/m2) | ??? | ??? | ??? | ??? |
| Systolic Blood Pressure (mmHg) | ??? | ??? | ??? | ??? |
| Diastolic Blood Pressure (mmHg) | ??? | ??? | ??? | ??? |
| Total Cholesterol (mg/dL) | ??? | ??? | ??? | ??? |
| Ideal Smoking Health | ??? | ??? | ??? | ??? |
| Ideal Physical Activity Health | ??? | ??? | ??? | ??? |
| Ideal Nutrition Health | ??? | ??? | ??? | ??? |
|  |  |  |  |  |
| Simple 6 | ??? | ??? | NA | NA |

\*The OR for each exposure is adjusted for all of the other exposure variables listed in the table (except Simple 6).

**Table 3.** **Exploration of potential total cholesterol by sex and BMI by sex interaction with fasting plasma glucose.**

|  |  |  |
| --- | --- | --- |
| Exposure | **Outcome = Fasting Plasma Glucose excluding diabetic individuals at visit 3** | |
|  | **β estimate ± SE** | **P-value** |
| Model 1 |  |  |
| Age (years) | ??? | ??? |
| Sex | ??? | ??? |
| Body mass index (kg/m2) | ??? | ??? |
| Total Cholesterol (mg/dL) | ??? | ??? |
| Total Cholesterol x Sex | ??? | ??? |
|  |  |  |
| Model 2 |  |  |
| Age (years) | ??? | ??? |
| Sex | ??? | ??? |
| Body mass index (kg/m2) | ??? | ??? |
| Total Cholesterol (mg/dL) | ??? | ??? |
| Body Mass Index x Sex | ??? | ??? |

**PASTE YOUR R CODE FOR PARTS 1-3 HERE.**