**EPI202: Fall 2022**

**Homework 4**

**To be uploaded as PDF to course website by 9:30 am on December 8, 2022**

Please provide brief but precise answers.

We encourage collaborative learning in this course. You may discuss homework assignments with other students. However, all written work that you submit for grading must be your own, in your own words, reflecting your understanding of the homework assignment. Homework assignments should not be prepared by copying, paraphrasing, or summarizing someone else’s work.

**Proper notation should be used throughout the assignment and relevant calculations should be shown.**

**Part I. Exploring the association between sedentary lifestyle and all-cause mortality using logistic regression models.**

In this section, you will use data from a prospective cohort study based upon the first 10 years of follow-up of participants in the Myocardial Infarction Onset Study to examine the relationship between sedentary lifestyle defined as physical activity less than once per week compared to one or more times per week (at baseline) and the cumulative incidence of death from any cause within the first 10 years. The variables in this dataset are described below. [Note: In this study, conducted in the late 1980s, sex was recorded and categorized as female or male based on NIH reporting requirements in place at the time. No data on gender was recorded]:

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| Id | ID number |
| Age | Age (continuous, years) |
| age\_cat | Age Category (1: <50yrs, 2: 50-64 yrs, 3: 65+ yrs) |
| Female | Female (1: female, 0: male) |
| Married | Married (1: yes, 0:no) |
| Educ | Educational Attainment (1: <HS, 2: HS, 3: >HS) |
| Dm | Diabetes (1: yes, 0:no) |
| Htn | Hypertension (1: yes, 0:no) |
| phys\_activity | Frequency of Physical Activity (0: <1/wk, 1: 1-3/wk, 2: 4+/wk) |
| follow\_up | Duration of follow-up (years) |
| Dead | Death from any cause in the first 10 years (0: Alive, 1: Died) |

*In homework 4, you will continue your analysis of the association between sedentary lifetyle at baseline and the outcome of all-cause mortality by ten years that you began with tabular methods last week in homework 3.*

You may use SAS, STATA, R or any other statistical analysis software package of your choosing. Include relevant lines of code. If you do any calculations by hand or using the Epi 202 calculator, identify the formulas you used and the values you included in the formulas, as in previous homework assignments.

1. In the last question of Homework 3, you evaluated the association between sedentary lifestyle (physical activity < 1 time per week versus ≥ 1 time per week) and all-cause mortality after adjusting for age using tabular analysis (i.e., Mantel-Hanszel estimator). However, the association might be confounded by many other variables. Here, you will analyze the association between sedentary lifestyle at baseline and 10-year all-cause mortality using logistic regression models and adjusting for other potential confounders**.** 
   1. **Write out and interpret the algebraic form** of the logistic regression model that includes as covariates: sedentary lifestyle (exposure), diabetes, hypertension, sex (female), and age (as a continuous variable).
      1. Write out the model

Text

Description automatically generated

* + 1. Interpret in words each coefficient term in your model.

The log odds of 10-year all-cause mortality is 5.76 among those who did not have a sedentary lifestyle, diabetes, hypertension, and aged 0, assuming no selection bias or information bias.

The log odds of 10-year all-cause mortality is lower among those who had a sedentary lifestyle, adjusted for diabetes, hypertension, sex and age, assuming no confounding, selection bias, or information bias.

The log odds of 10-year all-cause mortality is 0.850 lower among those who had diabetes, adjusted for sedentary lifestyle, hypertension, sex, and age, assuming no confounding, selection bias, or information bias.

The log odds of 10-year all-cause mortality is 0.341 lower among those who had hypertension, adjusted for sedentary lifestyle, diabetes, sex, and age, assuming no confounding, selection bias, or information bias.

The log odds of 10-year all-cause mortality is 0.195 higher among those who were female, adjusted for sedentary lifestyle, diabetes, hypertension, and age, assuming no confounding, selection bias, or information bias.

The log odds of 10-year all-cause mortality will decrease 0.064 per 1 year increase in age, adjusted for sedentary lifestyle, diabetes, hypertension, and sex, assuming no confounding, selection bias, or information bias.

* 1. **Expand your written model from (a)** to also include the interaction between a sedentary lifestyle and diabetes at baseline.

A picture containing text

Description automatically generated

* + 1. Write out the model

* + 1. Interpret the coefficient terms for: (a) sedentary lifestyle, (b) diabetes, and (c) the sedentary\*diabetes interaction term.

The log odds of 10-year all-cause mortality is 0.562 lower among those who did not have diabetes comparing having a sedentary lifestyle at baseline to not having a sedentary lifestyle at baseline, holding hypertension, sex, and age constant, assuming no confounding, selection bias, or information bias.

The log odds of 10-year all-cause mortality is 0.747 lower among those who did not have a sedentary lifestyle at baseline comparing having diabetes to not having diabetes, holding hypertension, sex, and age constant, assuming no confounding, selection bias, or information bias.

The difference between having diabetes and not having diabetes in log odds ratios for sedentary lifestyle at baseline and 10-year all-cause, holding hypertension, sex, and age constant, assuming no confounding, selection bias, or information bias.

* + 1. What is the association between sedentary lifestyle and death during the follow-up period among those with and without diabetes separately? Answer this question only using model coefficients. Report associations as odds ratios. **There is no need to interpret the results.**

Among those having diabetes,

Among those having no diabetes,

1. Next, you will fit the models you described above using statistical software.

[Hint: you can find useful code and examples in the regression guides posted on canvas]

* 1. Prepare Table 3 for your publication as follows.

|  |  |
| --- | --- |
| **Association between sedentary lifestyle and all-cause mortality** | OR  (95% CI) |
| Crude |  |
| Age and sex adjusted |  |
| Fully adjusted\* |  |

\*adjusted for…

* 1. From the fully adjusted model, estimate and interpret the odds ratio and 95% confidence interval for the association between a **5-year** increment in age and death during the follow-up period.
  2. Include an interaction term between sedentary lifestyle and **diabetes in your fully adjusted model.**
     1. Provide the adjusted odds ratio and 95% confidence interval for the association between sedentary lifestyle at baseline and death from any cause **among those with diabetes**. **Interpret your numerical result in words.**
     2. Provide the adjusted odds ratio and 95% confidence interval for the association between sedentary lifestyle at baseline and death from any cause **among those without diabetes**. **Interpret your numerical result in words.**
  3. Is there evidence for effect measure modification for the association between sedentary lifestyle by diabetes **on the multiplicative** scale in these data? Provide the p-value for the test of homogeneity and interpret it (i) directly and (ii) in the context of a Neyman-Pearson hypothesis test.
  4. After adjusting for diabetes, hypertension, sex (female), and age (as a continuous variable), is there evidence that the association between sedentary lifestyle at baseline and death from any cause is modified by diabetes **on the additive scale**?

(**HINT**: create a model recoding sedentary lifestyle and diabetes into a single four-level categorical variable and use the model output to calculate the relative excess risk due to interaction). 🡪 RERI = OR (sedentary=1, diabetes=1) – OR(sedentary =1, diabetes =0) – OR(sedentary =0, diabetes =1) + 1

1. Start to write the results section of a paper describing the relationship between sedentary lifestyle at baseline and death in the subsequent 10 years. Start by describing the study population. Describe the key findings from Table 1 (created as a part of homework 3), highlighting important differences between participants who were and were not sedentary at baseline, but don't repeat every number from the table. Rather, you should describe the results to give the reader insight into the study population and major differences between those with and without sedentary lifestyle at baseline. (**HINT**: You might want to look at some of the papers that you read for EPI201 and EPI202 throughout the semester for guidance as to the style of the first paragraph of the results section of an epidemiologic research paper.) For the next paragraph, describe the results from Table 2 (also created as part of homework 3). You should include information regarding the crude association and whether there appears to be confounding and/or effect modification by diabetes in the study data. In the third paragraph, summarize the findings from Table 3 (created as part of homework 4). Now describe the results of your evaluation of whether there is evidence for effect measure modification by diabetes on the multiplicative and additive scales in the fully adjusted model. Finally, conclude with a brief summary (two or three sentences) describing your conclusions based on these results.