**Analysis Exercise 2** (version April 13, 2020)

**Due via Canvas April 24, 5pm.**

Assignment is graded out of 100 points and is worth 15% of the MSCR final 534 grade.

Two files to be turned in:

1. This completed Word document: Answers to Questions 1-3, Completed Tables 1-3
2. Full SAS code (not SAS output); this is in addition to specific parts of Questions 1-3

**Note**:

* Students may work with classmates on the exercise, but each student must submit his or her own assignment.
* All relevant SAS codes and data needed to complete the assignment are in the Analysis Exercise assignment page in Canvas. Additional supporting documentation will be in lab and lecture files.
* Complete answers and tables using RED FONT.

*Assignment Purpose:*

1. To design and create tables for presentation of epidemiologic data.
2. To perform Poission regression, regression with splines, and use competing risk analysis with Cox Hazards Regression.
3. To generate and interpret measures of association using SAS.

**Question 1. Poisson Regression**

*For Question 1, use the SAS program code entitled “AnalysisExercise2Question1” on Canvas to run the datalines statement (which will create the dataset), then complete Questions 1 and Table 1 below.*

A researcher aimed to compare the automobile mortality rates among drivers aged 18-70 between two states (state 1 and state 0) during the year 2003. The total number count of deaths (see variable ‘deaths’) due to automobile accidents within each of eight age groups are listed in the datelines. The age groups are 18-22, 23-35, 36-41, 42-47, 48-53, 54-58, 59-63, and 64-70. The average number of people in each age group are included in the variable ‘population’ which was determined by the states’ Department of Motor Vehicles database on number of drivers’ licenses.

*Fill in Table 1. Use proc genmod to obtain the rate ratios and 95% confidence intervals.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 1.** | | | | | |
| **State** | **Total**  **Deaths** | **Total**  **Population** | **Crude**  **Rate (per 100,000)** | **Rate Ratio**  **(95%CI)** | **Adjusted Rate Ratio (95%CI)** |
| State 1 |  |  |  |  |  |
| State 0 |  |  |  | *REF* | *REF* |

*Question 1A. What is the interpretation of the adjusted rate ratio in Table 1?*

*Question 1B. What variable was used to indicate the offset in calculating the rate ratios in proc genmod? What information does this variable include and why is it needed?*

*Question 1C. Paste the SAS code here for the adjusted rate ratio calculation in Table 1.*

**Question 2. Splines**

*For Question 2, use the SAS dataset ‘cohort’ from Lab 8. In Lab 8 we used four equal cut-points and a quadratic spline in a Cox Hazards regression using a SAS macro. For Question 2, redo this spline regression analysis by using the same SAS macro with 3 equal knots (tertiles). Use the data to complete Table 2 and answer Questions 2 below.*

Details on the study design, SAS macro, and data dictionary are equivalent to Lab 8 (April 1, 2020).

*Question 2A. What were the cut points used for the tertiles of random blood glucose?*

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 2** | | | |
|  | | **Conventional HR** | |
| Tertile | RBG1 mg/dl | **HR2 (95%CI)** | **aHR3 (95%CI)** |
| 3 |  |  |  |
| 2 |  |  |  |
| 1 |  | REF | REF |
|  |  | **Quadratic spline with 3 equal knots4** | |
|  |  | **HR2 (95%CI)** | **aHR3 (95%CI)** |
| 3 |  |  |  |
| 2 |  |  |  |
| 1 |  | REF | REF |
| 1. Random blood glucose (RBG): Report three categorical levels of the random blood glucose based on tertile cut-points. Choose the lowest RBG level as the referent category 2. Crude unadjusted hazard ratio 3. Adjusted hazard ratio, adjusted for age category 4. Generated using SAS macro in Lab 8 | | | |

*Question 2B. Summarize the findings between the aHRs from the conventional regression and the models that used the quadratic spline? What additional information does the spline regression results suggest?*

*Question 2C. Paste the SAS code from the conventional and spline proc phreg procedures here. Do not paste the SAS macro.*

**Question 3. Competing risk analysis**

*For Question 3, use the SAS dataset entitled “AnalysisExercise2Question3” on Canvas. Use the data to complete Table 3A and Table 3B, and complete Questions 3 below.*

A cohort study aimed to determine the hazard rate of incident of invasive pulmonary Aspergillosis (fungal infection) in patients with idiopathic pulmonary fibrosis. The primary exposure of interest was biologic sex and a competing risk of concern was all-cause mortality. Assume all participants were free of Aspergillosis at baseline and were at risk of disease.

Data from the prospective cohort study included study visits every 6 months and variables from the study are listed below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Variable Description** | **Value** | **Label** |
| sex | Biologic sex | 0  1 | Female  Male |
| Censor1 | Censor variable with competing risks | 0 | No Aspergillosis/no death |
|  |  | 1  2 | Incident Aspergillosis  Death (all-cause) |
| age | Age group | 1 | 45 - 64 |
|  |  | 2 | 65 - 74 |
|  |  | 3  4 | 75 - 84  ≥ 85 |
| BMI | Obese | 0 | BMI <= 30 |
|  |  | 1  9 | BMI >30  Missing/unknown |
| Smk | Current smoking status | 0 | Non smoker/past smoker |
|  |  | 1  9 | Current smoker  Missing/unknown |
| Timetoc2 | Time to censor coding (continuous days) |  |  |

*Question 3A. What is the crude risk ratio for incident Aspergillosis comparing men to women? (Hint: this can be calculated from a 2 X 2 or by a log binomial model). For this answer, ignore competing risk due to death (e.g., include those who died in the denominator at risk).*

*Question 3B. What was the median follow up time for participants in this study?*

|  |  |  |
| --- | --- | --- |
| **Table 3A.** | | |
|  | **Conventional HR** | |
|  | HR1 (95%CI) | aHR2 (95%CI) |
| Male |  |  |
| Female | REF | REF |
| 1. Crude hazard ratio 2. Adjusted hazard ratio, adjusted for age and smoking status. | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3B.** | | | | |
|  | **Cause Specific** | | **Sub-distribution** | |
| Outcome: | Incident Aspergillosis | | Incident Aspergillosis | |
|  | *HR1 (95%CI)* | *aHR2 (95%CI)* | *HR1 (95%CI)* | *aHR2 (95%CI)* |
| Male |  |  |  |  |
| Female | REF | REF | REF | REF |
| Outcome: | Death | | Death | |
|  | *HR1 (95%CI)* | *aHR2 (95%CI)* | *HR1 (95%CI)* | *aHR2 (95%CI)* |
| Male |  |  |  |  |
| Female | REF | REF | REF | REF |
| 1. Crude hazard ratio 2. Adjusted hazard ratio, adjusted for age and smoking status. | | | | |

*Question 3C. From the sub-distribution model with outcome death, report the findings from the SAS output for “Summary of Failure Outcomes.”*

*Question 3D. Summarize the findings and conclusions based on Tables 3A and 3B.*

*Question 3E. Paste SAS code here used to generate Tables 3A and 3B.*