STAT 796: Homework 6

Due **Monday**, March 11 at 11:59pm on Canvas.

For questions 1 and 2 on this assignment, use data from the Burn Study. These data are described in HL and are available on Canvas in the file `burn.csv`.

1. Explore the relationship between total burn surface area (`tbsa`, in percentage point 0-100) and survival (coded as binary variable `death`).
    a. Provide a joint summary of total burn surface area and survival in this cohort, using a plot and/or descriptive statistics. What trends do you observe about their relationship in the data?
    b. Provide a joint summary of age (`age`, in years) and survival in this cohort, using a plot and/or descriptive statistics. What trends do you observe about their relationship in the data?
    c. Fit a logistic regression model with total burn surface area as the predictor of interest and death as the outcome. Adjust for indicator of flame being involved in the injury (binary variable `flame`), age, and indicator of white race (binary variable `white`). Provide a point estimate, confidence interval, and interpretation for the odds ratio corresponding to total burn surface area.
    d. Make two plots that show the estimated relationship between log odds of death and (i) total burn surface area and (ii) age from the model in (c). Log odds of death should be the vertical axis and the predictor variable on the horizontal axis.
    e. Fit the model from (c), but now including age using natural splines with 3 df. Provide a point estimate, confidence interval, and interpretation for the odds ratio corresponding to total burn surface area.
    f. Make a plot showing the estimated relationship between log odds of death and age from the model in (e).
    g. Which model fits the data better? Support your answer with the results of a hypothesis test.
    h. Are your results about the relationship between total burn surface area and survival substantively different between (c) and (e)?
    i. Fit the same model as (e), but now use natural splines for total burn surface area. Make a plot of the estimated relationship between log odds of death and total burn surface area from this model. How does this compare to your summary in (a)?
2. Explore the evidence for effect modification of the relationship between total burn surface area and survival.
    a. Fit a logistic regression model with total burn surface area as the predictor of interest and death as the outcome. Include an interaction between total burn surface area and indicator of flame being involved in the injury. Also adjust for age (either as a linear term or using splines, based on your response to Question 1) and indicator of white race. Provide a point estimate for odds ratios corresponding to total burn area among people with and without flame involved in their burn injury.
    b. Is there evidence that the odds ratios in (a) are different from one another? Support your answer with the results of a hypothesis test.
3. Choose a research question from your work (dissertation research, class project, etc.).
    a. State the exposure of interest and the outcome of interest.
    b. Provide two relevant variables that are confounders, and explain the relationships that make them confounders in this setting.
    c. Provide two variables related to the exposure and/or outcome that are *not* confounders, and explain the relationships and why they are not confounders in this setting.