STAT 796: Homework 3

Due Sunday, February 10 at 11:59pm on Canvas.

For this assignment, you will analyze data from the "Myopia" dataset described in HL. These data come from the Orinda Longitudinal Study of Myopia and the subsequent Collaborative Longitudinal Evaluation of Ethnicity and Refractive Error (CLEERE) Study, which were designed to investigate risk factors for myopia ("nearsightedness") in children. This dataset is available in the file `myopia.csv` on Canvas, and a description of its coding is also provided there.

Please provide your code in an Appendix at the end of your responses.

1. The following questions ask about descriptive summaries of the cohort.
    a. What proportion of the cohort had myopia within 5 years of study entry?
    b. Report the proportion of children whose mom had myopia among (i) all children, (ii) children without myopia, and (iii) children with myopia.
    c. Report the proportion of children whose dad had myopia among (i) all children, (ii) children without myopia, and (iii) children with myopia.
    d. Report the proportion of children who were male among (i) all children, (ii) children without myopia, and (iii) children with myopia.
    e. Make a boxplot of the hours of TV watched by children, stratified by myopia status.
    f. Make a boxplot of the hours of computer time for the children, stratified by myopia status.

2. For this question, fit a logistic regression model using myopia status as the outcome. Include age, indicator of sex, indicator of mom with myopia, indicator of dad with myopia, hours spent watching TV, and hours spent on the computer as the predictor variables.
    a. For each of the three indicator predictor variables, report the coefficient estimate and interpret its meaning in a single sentence.
    b. For each of the three indicator predictor variables, report an estimated odds ratio and interpret its meaning in a single sentence.
    c. How do the odds ratios you calculated in (b) compare to the data summaries in (b)-(d) of Question 1?
    d. For each of the three other (non-indicator) predictor variables, report an estimated odds ratio and interpret its meaning in a single sentence.
    e. What is the estimated adjusted odds ratio for myopia corresponding to a difference of 5 hours in amount of TV watched each week?
    f. What is the estimated adjusted odds ratio for myopia corresponding to a difference of 25 hours in amount of computer time each week?
    g. What is the estimated adjusted odds ratio for myopia corresponding to a difference of 5 years in age at first visit?
    h. Do the quantities calculated in (e)-(g) make sense in the context of this cohort? *[This is a subjective question, but consider the range of these predictor variables in the cohort.]*