Biostatistics 515/518, Winter 2020

HW5 (4 problems)

Questions 1-3 use the MRI dataset used in previous assignments. The file "mri.txt" contains the data and the documentation is in the file "mri.pdf"

1. We are interested in examining how mean systolic blood pressure varies by age and sex.
    a. Create a scatterplot of systolic blood pressure versus age. Use different symbols and/or colors for each **sex group**, and include LOWESS (or LOESS) smooths for each sex group separately.
    b. Do you see evidence from the scatterplot that sex modifies the association between systolic blood pressure and age? Explain your reasoning.
    c. Perform a formal statistical analysis to investigate whether sex modifies the association between systolic blood pressure and age. Provide full statistical inference and reporting of your findings.
    d. Regardless of statistical significance in part c, interpret the model you fit in part c. Do results make sense in light of the plot you made in part a? Comment.

2. We now turn our attention to systolic blood pressure and age, with race as the covariate.
    a. Write the form of a multiple regression model for examining the race-adjusted association between mean systolic blood pressure (as the response) and age. Make sure the definitions of variables you include in your model and variable coding are clear.
    b. Fit your model to the dataset and interpret the results (age is the predictor of interest). Use language suitable for a scientific publication, and include an interpretation of any confidence interval.

3. .
    a. Fit a multiple regression model to predict systolic blood pressure based on patient age, sex, and race. Write out your fitted model.
    b. Give a 95% prediction interval for systolic blood pressure among 70-year-old black women.
    c. Make a plot of residuals vs. fitted values and a QQ plot of the residuals. Comment on whether cast doubt on any of the assumptions required for the validity of prediction. (Note: the course may not cover these diagnostics before this assignment is due, so 3c may become optional.)

4. Consider the results from a study, which may have been a cross-sectional study, a cohort study, or a case-control study. D and E are indicator variables for "disease" and "exposure," respectively.

	D=0	D=1	
E=0	a	b	a+b
E=1	c	d	c+d
	a+c	b+d	n=a+b+c+d

a. Derive the odds ratio for disease given exposure.
b. Derive the odds ratio for exposure given disease.
c. Compare the two odds ratios in A and B.

NOTE: in a and b you should start from the definitions of odds and take ratios.