
BIOS 507 HOMEWORK 6

Due 4/7/2025 by 11:59pm

Directions: Complete all questions. Any R or SAS code used should be attached at the end of the homework. Collaboration is encouraged, but the final product must be your own work.

Problem 1

A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 15 women from each 10-year age group, beginning with age 40 and ending with age 79. The dataset is "musclemass.txt". Note: These questions are adapted from questions 8.4 and 8.5 in the textbook.

Part A

Fit a quadratic regression model with centered age. Plot the fitted regression function and the data. Does the quadratic regression function visually appear to be a good fit here? Report the R^2 value.

Part B

Conduct an overall test for model fit using $\alpha = 0.05$. What are your findings?

Part C

Test whether the quadratic term can be dropped from the model using $\alpha = 0.05$. What are your findings?

Part D

Find and interpret a 95% confidence interval for the mean muscle mass for a women who is age 50.

Part E

Find and interpret a 95% prediction interval for the muscle mass for a women who is age 50.

Part F

Fit the third-order model and test for the significance of the cubic term (using $\alpha = 0.05$).

Problem 2

A researcher is interested in understanding the relationship between income (measured in thousands of dollars per year) and the amount of money spent yearly on food (also measured in thousands of dollars). The researcher has collected data on these two variables, which can be found on Canvas under `food_income.csv`. Conduct an analysis to model the relationship between these two variables, using the food expenditures the outcome. Be sure to state all assumptions, check their validity, describe any statistical tests you perform, and perform any required remedial measures.

Problem 3

In a previous homework, you analyzed a random sample of data on infants' birth weights (Y , lbs) and gestation period (X_1 , weeks). I've uploaded a new version of this data to canvas called `birth_weight_expanded.csv`. There are 4 variables in this data set:

- `SNAP Score for Neonatal Acute Physiology`, a continuous measure of illness severity. Higher scores indicate more severe illness.
- `birthweight` Birth weight in pounds.
- `gestation period` gestational age at birth, in weeks.
- `sex` whether the infant is male or female

The researcher is interested in understanding whether birth weight and/or gestation period are useful for predicting the SNAP score. The researcher also wants to be sure to control for sex. Conduct regression analyses to answer the researcher's question. Be sure to list and check your assumptions, describe any hypothesis testing you perform (including full and reduced models), and describe any remedial measures used.