PART 2 (20 points)

The data set kid.weights (Library: UsingR) contains age, weight, and height measurements for several children.
(a) Fit the linear model shown below and report summary of the model.
$$weight = b_0 + b_1 age + b_2 height + b_3 height^2 + b_4 height^3 + b_5 height^4$$

(b) Use the partial F-test to select between this model and the nested models found by using only first-, second-, and third-degree polynomials for height.

PART 3 (20 points)

(Use TermLife.csv data file) Term Life Insurance: Here we examine the 2004 Survey of Consumer Finances (SCF), a nationally representative sample that contains extensive information on assets, liabilities, income, and demographic characteristics of those sampled (potential U.S. customers). We study a random sample of 500 families with positive incomes. From the sample of 500, we initially consider a subsample of n = 275 families that purchased term life insurance.

Note: For n = 275, we want you to subset the data so that you are only looking at rows where FACE > 0. Also, variable LNFACE = log of the face variable and LNINCOME = log of the income variable.

(a) Fit a linear regression model of LNINCOME, EDUCATION, NUMHH, MARSTAT, AGE, and GENDER on LNFACE.

(b) Check if multicollinearity is present.

(c) Briefly explain the idea of collinearity and a variance inflation factor. What constitutes a large variance inflation factor?

(d) Supplement the variance inflation factor statistics with a table of correlations of explanatory variables. Given these statistics, is collinearity an issue with this fitted model? Why or why not?

PART 4 (20 points)

(Use condo.csv data file) A real estate agent wishes to determine the selling price of residences using the size (square feet) and whether the residence is a condominium or a single- family home.

(a) Fit a regression model to predict the selling price for residences and provide the regression equation.

(b) Interpret the parameters β_1 and β_2 in the model given in part (a).

(c) Fit a new regression model now including the interaction term x1* x2 and provide the regression equation.

(d) Describe what including this interaction term accomplishes.

(e) Conduct a test of hypothesis to determine if the relationship between the selling price and the square footage is different between condominiums and single-family homes.

PART 5 (20 points)

The data set fat (Library: UsingR) contains several body measurements that can be done using a scale and a tape measure. These can be used to predict the body-fat percentage (body.fat). Measuring body fat requires a special apparatus; if our resulting model fits well, we have a low-cost alternative.

(a) Partition the data into 60% for training and 40% for testing. Use set.seed(25) before data partition.

(b) Use training data to develop a multiple linear regression model with body.fat as response variable and age, weight, height, BMI, neck, chest, abdomen, hip, thigh, knee, ankle, bicep, forearm, and wrist as independent variables.

(c) Use the stepAIC function to select a model. Report model summary and provide equation for this model.

(d) What are the top three contributors to the body-fat percentage? Provide an interpretation for these three coefficients.

(e) Develop a scatter plot for predicted and fitted response values using the testing data. Obtain R^2 using testing data based on predicted and fitted response values?