Regression Analysis

Assignment 3B

Problem 4.5 on page 185: This exercise does not require use of statistical software. It covers predictor interactions (pages 159-166).

Problem 4.5: This problem extends the home prices example used previously to 76 homes (Section 6.1 contains a complete case study of these data). We wish to model the association between the price of a single-family home (*Price* in $ thousands) and the following predictors:

Floor = floor size (thousands of square feet)
Lot = lot size category (from 1 to 11—see page 89)
Bath = number of bathrooms (with half-bathrooms counting as "0.1")
Bed = number of bedrooms (between 2 and 6)
Age = age (standardized: (year built –1970)/10)
Gar = garage size (0, 1, 2, or 3 cars)
D_{Ac} = indicator for "active listing" (rather than pending or sold)
D_{Ed} = indicator for proximity to Edison Elementary
D_{Ha} = indicator for proximity to Harris Elementary

Consider the following model, which includes an interaction between *Bath* and *Bed*:

$E(Price) = b_0 + b_1 Floor + b_2 Lot + b_3 Bath + b_4 Bed + b_5 BathBed + b_6 Age + b_7 Age^2 + b_8 Gar + b_9 D_{Ac} + b_{10} D_{Ed} + b_{11} D_{Ha}$.

The regression results for this model are:

Predictor variable	Parameter estimate	Two tail p-value
Intercept	332.47	0.00
Floor	56.72	0.05
Lot	9.92	0.01
Bath	-98.15	0.02
Bed	-78.91	0.01
BathBed	30.39	0.01
Age	3.30	0.30
Age^2	1.64	0.03
Gar	13.12	0.12
D_{Ac}	27.43	0.02
D_{Ed}	67.06	0.00
D_{Ha}	47.27	0.00

Hint: Understanding the ***SALES1*** *example beginning on page 159 will help you solve this problem.*

a. Test whether the linear association between home price (*Price*) and number of bathrooms (*Bath*) depends on number of bedrooms (*Bed*), all else equal (use significance level 5%).
b. Does the linear association between *Price* and *Bath* vary with *Bed*? We can investigate this by isolating the part of the model involving just *Bath*: the "*Bath* effect" on *Price* is given by $b_3 Bath + b_5 BathBed = (b_3 + b_5 Bed) Bath$. For example, when *Bed* = 2, this effect is estimated to be $(-98.15 + 30.39(2)) Bath = -37.37 Bath$. Thus, for two-bedroom homes, there is a negative linear association between home price and number of bathrooms (for each additional bathroom, the sale price drops by $37,370, all else being equal—perhaps adding extra bathrooms to two-bedroom homes is considered a

waste of space and so has a negative impact on price). Use similar calculations to show the linear association between *Price* and *Bath* for three-bedroom homes, and also for four-bedroom homes.

[4 points]

Problem 4.7 on page 187: This exercise uses the HOMES5 dataset and covers qualitative predictors with two levels (pages 166-174). Answering this exercise should be relatively straightforward if you fully understand Section 4.3.1; otherwise it might seem a little mysterious. The equation that part (a) asks you to write down is intended to have regression parameters (specifically b_0, b_1, b_2, and b_3), and the predictor terms D_{Ha}, *Floor*, and $D_{Ha}Floor$. The equations that part (e) asks you to write down are intended to have regression parameter estimates (i.e., actual numbers), and just the predictor term *Floor*.

Problem 4.7: Consider the data available in the **HOMES5** data file—these data are for 40 single-family homes in south Eugene, Oregon in 2005 and were provided by Victoria Whitman, a realtor in Eugene. For the sake of illustration, here we investigate whether any linear association between *Floor* (floor size in thousands of square feet) and *Price* (sale price in thousands of dollars) differs for two particular neighborhoods (defined by closest elementary school) in this housing market—we ignore the other predictors (lot size, age, etc.) for now. (We'll revisit this application in Section 6.1 when we consider a complete analysis of the whole dataset.) In the **HOMES5** data file there are 26 homes whose closest elementary school is "Redwood" ($D_{Ha} = 0$) and 14 homes whose closest elementary school is "Harris" ($D_{Ha} = 1$).

*Hint: Understanding the **SALGPA2** example beginning on page 171 will help you solve this problem.*

a. Write the equation of a model relating sale price (*Price*) to floor size (*Floor*) and neighborhood (D_{Ha}) that allows for different slopes and intercepts for each neighborhood.

   Hint: You should have regression parameters (b's) and predictor terms in your equation, but no numbers.

b. Draw a scatterplot that illustrates the model in part (a). Include two regression lines, one for each neighborhood, on your plot [computer help #15, #17, and #33].
c. Use statistical software to fit the model from part (a) to the data and write out the resulting estimated regression equation. You will first need to create the "$D_{Ha}Floor$" interaction term in the dataset [computer help #6 and #31].

   Hint: This is where you replace the regression parameters in your equation from part (a) with numbers obtained using statistical software.

d. Conduct a hypothesis test to determine whether the slopes associated with the two neighborhoods are significantly different. Use significance level 5%.
e. Based on the results from part (d), fit a new model to the data, and write two separate equations (with actual numbers) for predicting *Price* from *Floor*, one for the Redwood neighborhood and the other for the Harris neighborhood.

[10 points]