Stat 960:587 Homework No. 1
Due 2/20/2019

Programming: R or Python

Problem 1. The following data (Natrella, 1963) are on the latent heat of the fusion of ice *(cal/gm):*

Method A: 79.98 80.04 80.02 80.04 80.03 80.03 80.04 79.97
80.05 80.03 80.02 80.00 80.02
Method B: 80.02 79.94 79.98 79.97 79.97 80.03 79.95 79.97

(a) Assuming normality, test the hypothesis of equal means, both with and without making the assumption of equal variances.
(b) Inspect the data graphically in various ways, for example, boxplots, Q-Q plots and histograms.
(c) Pool the data together and plot a estimated density function of the pooled data. Use the triangle kernel with bandwidth h = 0.015 to provide an estimate of the density function f(x) at x = 80. What is your estimate and provide your calculation.

Problem 2: Brand preference. In a small-scale experimental study of the relation between degree of brand liking (Y) and moisture content $(X_1)$and sweetness (X_2) of the product, the following results were obtained from the experiment based on a completely randomized design (data are coded):

i:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X_1:	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10
X_2:	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Y_i:	64	73	61	76	72	80	71	83	83	89	86	93	88	95	94	100

a. Fit two regression models to the data, one with main covariate terms plus the covariate interaction term and the other with main covariate terms but without the interaction term. State the estimated regression functions. How is $\hat{\beta}_1$ interpreted here?
b. Obtain the residuals and prepare a box plot of the residuals for the model with both covariates and their interaction. What information does this plot provide?
c. Plot the residuals against $\hat{Y}, X_1, X_2,$ and X_1X_2 on separate graphs. Analyze the plots and summarize your findings.
d. Conduct a formal test for lack of fit of the first-order regression function; use a = .01. State the alternatives, decision rule, and conclusion.

Problem 3: Refer to **Brand preference.** The diagonal elements of the hat matrix are: $h_{55} = h_{66} = h_{77} = h_{88} = h_{99} = h_{10,10} = h_{11,11} = h_{12,12} = .137$ and $h_{11} = h_{22} = h_{33} = h_{44} = h_{13,13} = h_{14,14} = h_{15,15} = h_{16,16} = .237.$

a. Explain the reason for the pattern in the diagonal elements of the hat matrix.
b. According to the rule of thumb often stated in a regression text book, are any of the observations outlying with regard to their *X* values.
c. Case 14 appears to be a borderline outlying *Y* observation. Obtain the *DFFITS, DFBETAS,* and Cook's distance values for this case to assess its influence. What do you conclude?
d. Calculate the average absolute percent difference in the fitted values with and without case 14. What does this measure indicate about the influence of case 14?
e. Calculate Cook's distance *D*, for each case. Are any cases influential according to this measure?