STAT 3008
Exercise 8

Problems refer to the problem sets in the textbook: Applied Linear Regression, 3rd edition by Weisberg.

1. Problem 5.1.1, 5.1.2 and 5.1.3. **(Galtons sweet peas)**
   Many of the ideas of regression first appeared in the work of Sir Francis Galton on the inheritance of characteristics from one generation to the next. In a paper on "Typical Laws of Heredity," delivered to the Royal Institution on February 9, 1877, Galton discussed some experiments on sweet peas. By comparing the sweet peas produced by parent plants to those produced by offspring plants, he could observe inheritance from one generation to the next. Galton categorized parent plants according to the typical diameter of the peas they produced. For seven size classes from 0.15 to 0.21 inches, he arranged for each of nine of his friends to grow 10 plants from seed in each size class; however, two of the crops were total failures. A summary of Galtons data was later published by Karl Pearson (1930) (see Table 1 and the data file `galtonpeas.txt`). Only average diameter and standard deviation of the offspring peas are given by Pearson; sample sizes are unknown.

TABLE 1 Galtons Peas Data

Parent Diameter (.01 in)	*Progeny* Diameter (.01 in)	SD
21	17.26	1.988
20	17.07	1.938
19	16.37	1.896
18	16.40	2.037
17	16.13	1.654
16	16.17	1.594
15	15.98	1.763

   (a) Draw the scatterplot of *Progeny versus Parent*.

   (b) Assuming that the standard deviations given are population values, compute the weighted regression of *Progeny on Parent*. Draw the fitted mean function on your scatterplot.

(c) Galton wanted to know if characteristics of the parent plant such as size were passed on to the offspring plants. In fitting the regression, a parameter value of $\beta_1 = 1$ would correspond to perfect inheritance, while $\beta_1 < 1$ would suggest that the offspring are "reverting" toward "what may be roughly and perhaps fairly described as the average ancestral type" (The substitution of "regression" for "reversion" was probably due to Galton in 1885). Test the hypothesis that $\beta_1 = 1$ versus the alternative that $\beta_1 < 1$.

2. Problem 6.1.1, 6.1.2. **(Cake data)**
The data for this example are in the data file `cakes.txt`.

   (a) Fit (6.4) and verify that the significance levels are all less than 0.005.

   (b) Estimate the optimal (X_1, X_2) combination $(\tilde{X}_1, \tilde{X}_2)$ and find the standard errors of X_1 and X_2.

3. Problem 6.4.
The data in the file `twins.txt` give the IQ scores of identical twins, one raised in a foster home, IQf, and the other raised by birth parents, IQb. The data were published by Burt (1966), and their authenticity has been questioned. For purposes of this example, the twin pairs can be divided into three social classes C, low, middle or high, coded in the data file 1, 2, and 3, respectively, according to the social class of the birth parents. Treat IQf as the response and IQb as the predictor, with C as a factor.

   Perform an appropriate analysis of these data. Be sure to draw and discuss a relevant graph. Are the within-class mean functions straight lines? Are there class differences? If there are differences, what are they?

4. Problem 6.14.
Using the salary data in Problem 6.13, one fitted mean function is

   $$E(Salary|Sex, Year) = 18223 - 571Sex + 741Year + 169Sex \times Year$$

   (a) Give the coefficients in the estimated mean function if Sex were coded so males had the value 2 and females had the value 1 (the coding given to get the above mean function was 0 for males and 1 for females).

(b) Give the coefficients if *Sex* were codes as -1 for males and +1 for females.

5. Problem 6.15.1 and 6.15.3
   Pens of turkeys were grown with an identical diet, except that each pen was supplemented with an amount A of an amino acid methionine as a percentage of the total diet of the birds. The data in the file `turk0.txt` give the response average weight *Gain* in grams of all the turkeys in the pen for 35 pens of turkeys receiving various levels of A.

   (a) Draw the scatterplot of *Gain* versus A and summarize. In particular, does simple linear regression appear plausible?

   (b) To the graph drawn in (a), add the fitted mean functions based on both the simple linear regression mean function and the quadratic mean function, for values of A in the range from 0 to 0.60, and comment.