

R Homework for Chapter 19

Pizza ratings

Manufacturers of frozen foods often reformulate their products to maintain and increase customer satisfaction and sales. So they pay particular attention to evaluations of their products in comparison to their competitors' products. Frozen pizzas are a major sector of the frozen food market, accounting for an average of \$3.1 billion in sales in the five-year period from 2008 to 2012. The prestigious Consumer's Union rated frozen pizzas for flavor and quality, assigning an overall score to each brand tested. Let's build a regression model to predict the Consumer's Union score from *Calories*, *Type* (1 = cheese, 0 = pepperoni), and *Fat* content.

You can find the data file on Blackboard. Download it and put it in the **same folder** as your R program file. Then, use the following command to read the data

```
pizza <- read.table('Pizza_ratings.txt', sep = '\t', header = TRUE)
```

1. Regress *Score* on *Calories*, *Type* and *Fat*. Write down the fitted model. What is the interpretation of the coefficient of *Type* in this regression? Is that coefficient statistically significant? Explain.
2. Based on the model obtained above plot its residuals against predicted values. Do you see any unusual observations in the plot? If yes, please identify them and find their *standardized* residuals, leverages, Cook's distances and DFFITS measures. Are the unusual points reflected in the residual plot influential? Is there any other influential case you can find? If yes, please identify the corresponding pizza.
3. Let's remove *all* the influential cases from the dataset and refit the multiple regression model in Problem 1. Compare the new model to the old one based on their summaries. Check the assumptions for the new model.
4. Check collinearity for the model fitted in Problem 3. Does there exist any serious collinearity? If does, could you find the reason?
5. We now use the full dataset `pizza`. Do we need to consider the interaction between *Calories* and *Type*? Explain. Add an interaction term to the model if you think it is necessary, and fit the new model. Interpret the resulting coefficient of interaction term.