Problem 1
Tell true or false of the following statements and briefly explain your answer.

(a) To quantify a qualitative variable with three classes C_1, C_2, C_3, we need the following dummy variables:

$$X_1 = \begin{cases} 1 & \text{if } C_1 \\ 0 & \text{otherwise} \end{cases} \quad X_2 = \begin{cases} 1 & \text{if } C_2 \\ 0 & \text{otherwise} \end{cases} \quad X_3 = \begin{cases} 1 & \text{if } C_3 \\ 0 & \text{otherwise} \end{cases}$$

(b) Polynomial regression models with higher than the third power terms are preferred since they provide better approximations to the regression relation.

(c) In interaction regression models, the effect of one variable depends on the value of another variable with which it appears together in a cross-product term.

(d) With a qualitative variable, the best way is to fit separate regression models under each of its classes.

Solution:

(a) False. You can have a more compact representation with two variables for example.

(b) False. Likely to overfit.

(c) True. Interaction models are used to explain how variables interact with one another's values.

(d) False. You lose a lot of information if you don't allow the variables to be able to learn from one another.