2.

Conceptual

1. For each of parts (a) through (d), indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.

   (a) The sample size n is extremely large, and the number of predictors p is small.

   (b) The number of predictors p is extremely large, and the number of observations n is small.

   (c) The relationship between the predictors and response is highly non-linear.

   (d) The variance of the error terms, i.e. $\sigma^2 = \text{Var}(\epsilon)$, is extremely high.

Answers:

a) A flexible statistical learning method would perform better since the number of samples is extremely large. As a result, the model can't change its shape by a great deal with different data sets too.

b) A flexible statistical method would perform worse since the number of samples is small. The model can change its

shape/parameters by a great deal with a different data set. This is a classic example of over-fitting.

c) A flexible statistical method would perform better since the relationship between the predictors and response is non-linear. We cannot use an inflexible linear model in such scenarios. A flexible statistical method is best suited for non-linear relationships.

d) A flexible statistical method would perform worse as such a method would overfit the model by taking into consideration the high variance errors. Since the errors have a high variance, our model would conform/fit according to these errors, even if the predictor data doesn't have a high variance.