1. In order to do a regression model diagnostics, you need to do the following:

   ( ) Use numerical tools.
   ( ) Use graphical tools.
   (X) Use numerical and graphical tools.
   ( ) You do not need to worry about regression model diagnostics.
   Justification: We use a combination of statistical tests and plots to draw conclusions.

2. As a rule of thumb, an observation has a high leverage in the regression model if:

   ( ) The corresponding element in the diagonal of the **H** matrix, h_i is equal to 1.
   ( ) The Mahalanobis distance between the observation and the mean of all observations is greater than 1.
   (X) The corresponding element in the diagonal of the **H** matrix h_i is greater than 2×mean value of the diagonal elements $h_1, h_2, \ldots, h_n$.
   ( ) The variance of the predicted value $\hat{y}_i$ is greater than σ^2.
   Justification: We said that we check whether h_i is larger than $2p/n$.

3. In order to do an outlier test, we need to use one of the following:

   ( ) An F distribution
   ( ) A Chi-square distribution
   (X) A t distribution
   ( ) A Normal distribution
   Justification: Recall that we compare the value of the studentized residuals with a t distribution with $n - p - 1$ DFs.

4. According to Cook's distance, an observation can be highly influential in the regression model due to the following reasons (select all correct options):

   (X) The observation is an outlier.
   (X) The observation has a high leverage.
   (X) The observation is an outlier and has a high leverage.
   ( ) The observation does not have a Normal distribution.
   Justification: In the lectures we said that a highly influential point, according to Cook's distance, if it is an outlier, or if it is a high leverage point or both.

5. A student asked: "Why is it necessary to perform diagnostic checks of the fit when R^2 is large?" Comment.
   Justification: R^2 **only** tells us if the straight line is a good fit in the sense whether the predictors we have included in our model assuming a linear relationship with y help reduce the variation in y. However, if the model assumptions, of constant variance, normality, etc. are violated, then all our hypothesis tests will no longer be valid and in most cases, we will not be able to make inferences about the model parameters.