1. The members of a health spa pay annual membership of dues of \$300 plus a charge of \$2 for each visit to the spa. Let y denote the dollar cost for the year for a member and x the number of visits by the member this year.
   (a) The relation between x and y can be described as:
   ( ) $y_i = 300 + 2x_i + \varepsilon_i$ (X) $y = 300 + 2x$ ( ) None of the above.
   (b) Is it a functional relation or a statistical relation?
   (X) It is a functional relationship. ( ) It is a statistical relationship.

2. When asked to state the simple linear regression model, a student wrote it as follows: $E(y_i) = \beta_0 + \beta_1 x_i + \varepsilon_i$. Do you agree?
   ( ) Yes. (X) No.
   Justification: The ε term is a random variable with mean 0. So, the correct expression here would be: $E(y_i) = \beta_0 + \beta_1 x_i$ or $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

3. Evaluate the following statement: "*For the least-squares method to be fully valid, it is required that the distribution of y be normal.*"
   ( ) The above statement is correct. (X) The above statement is incorrect.
   Justification: Recall the derivation of the Least-Squares; we did not use the normality assumption anywhere. We just minimized the RSS with respect to the coefficients. An assumption regarding the distribution of the error terms is necessary when we want to derive statistical properties of the estimators, or when we want to do hypothesis testing etc.

4. According to our discussion, $\sum_i r_i = 0$ where r_i are the model residuals of a simple linear regression model fitted to a set of n cases using least-squares.
   Is it also true that $\sum_i \varepsilon_i = 0$
   ( ) Yes, it is correct that $\sum_i \varepsilon_i = 0$. (X) No, it is not correct that $\sum_i \varepsilon_i = 0$.
   Justification: The ε terms are random variables.

5. We want to fit a linear regression between speed (x) (mph) and the distance (y) taken to stop (ft) from measurements from a groups of cars. The following statistics are available from the 'cars' data set in the R datasets library:
   $S_{XX} = \sum_i (x_i - \bar{x})^2 = 1370$, $S_{YY} = \sum_i (y_i - \bar{y})^2 = 32538.98$, $r_{XY} = 0.807$.
   The estimated slope $\hat{\beta}_1$ is given by:
   (X) 3.93 ( ) 2.93 ( ) -3.93 ( ) 0.17
   Justification: We just use the formula for the LS estimator $\hat{\beta}_1$.

6. In the regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (as defined in Lecture 2), the error terms
   ( ) cannot be estimated. ( ) have a variance proportional to the mean. (X) have zero mean, constant variance and are independent. ( ) have unknown mean. ( ) none of the above.
   Justification: The ε terms are not parameters to be estimated. In the SLR model, we **assume** that the error terms have zero mean, constant variance and are independent.