1. A simple linear regression model was fitted to the `cars` data set from the **R** library data sets. For this model, the residual sum of squares is $RSS = \sum_i (y_i - \hat{y}_i)^2 = 11353.53$

   The total number of observations is $n = 50$. Then, the estimated residual variance $\hat{\sigma}^2$ is 236.53.
   Justification: The estimator for $\hat{\sigma}^2$ is $RSS/df_{RSS}$.

2. A simple linear regression model was fitted to the `cars` data set from the **R** library data sets, where $y$ represents the stopping distance (ft) and $x$ represents the speed (mph).

   If the Total variation in the response or total sum of squares is $TSS = 32538.98$ and the Residual Sum of Squares is $RSS = 11353.52$, the resulting coefficient of determination or $R$-square for the regression is 0.65
   Justification: $R^2 = 1 - RSS/TSS$.

3. A simple linear regression model that passes through the origin was fitted to the cars data set from the R library data sets, where y represents the stopping distance (ft) and x represents the speed (mph).

   If $\sum_i y_i^2 = 124903$ and $\sum_i \hat{y}_i^2 = 111949.2$ the value of the coefficient of determination for this regression model is: 0.896
   Justification: When we fit a regression through the origin, $\tilde{R}^2 = \frac{\sum_i \hat{y}_i^2}{\sum_i y_i^2}$.

4. For an observed value of speed $x_i$=14 mph, there is an observed value of stopping distance $y_i$=80 ft.

   According to the least-square estimated regression equation between distance and speed, the predicted stopping distance $\hat{y}_i = 37.47463$. The value of the estimated residual $r_i$ is given by one of the following options:

   ( ) 23.47463 ( ) -42.52537 ( ) -23.47463 (X) 42.52537
   Justification: The residulas are computed as $r_i = y_i - \hat{y}_i$.

5. The following output is obtained from fitting a simple linear regression to Galton's dataset: `![cars.png](/files/fa57ea0e-d5b1-41ae-a736-f455391d5290)` Select the correct simple linear regression fitted to this data.

   ( ) $y_i = 4.26511 + 0.06161x_i$ ( ) $y_i = 0.06161 + 4.26511x_i$ ( ) $y_i = 0.63736 + 22.63624x_i$ (X) $y_i = 22.63624 + 0.63736x_i$

6. In a study of the relationship for senior citizens between physical activity and frequency of colds, participants were asked to monitor their weekly time spent in exercise over a five-year period and the frequency of colds. the study demonstrated that a negative statistical relation exists between time spent in exercise and frequency of colds. The investigator concluded that increasing time spent in exercise is an efficient strategy for reducing the frequency of colds for senior citizens.

   (a) Were the data collected in the study observational or experimental data?

   (X) Observational ( ) Experimental Justification: There is no intervention from the experimenter, so this is an observational study.

(b) Comment on the validity of the conclusions reached by the investigator.

The investigator's statement is a causal statement, and such conclusions cannot be reached by an observational study.

(c) Identify two or three other explanatory variables that might affect both the time spent in exercise and the frequency of colds for senior citizens simultaneously. How might the study be changed so that a valid conclusion about causal relationship between amount of exercise and frequency of colds can be reached?

Explanatory variables: age, other medical conditions, time spent inside....
A randomized controlled experiment should be performed.