1. In the age of "big data", it is much more common for data scientists to specify their hypotheses *before* collecting data. 3 / 3 points

   True

   False

2. Circular data analyses, or double dipping, is: 4 / 4 points

   The process of exploring a dataset in an attempt to discover what relationships exist, and then test hypotheses related to that exploration on the same dataset.

   Something to be avoided.

   A process by which a data scientist can extract the most information from their dataset.

   The best way to conduct an exploratory data analysis.

3. Consider a dataset containing three variables, each measured in dollars: 4 / 4 point

   - Sales of a product (response, Y)
   - YouTube advertising budget (predictor, x_1)
   - Facebook advertising budget (predictor, x_2)

   Suppose that the relationship between the predictors and the response is linear, i.e., $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. What is the interpretation of β_2?

   β_2 is the average change in sales for a one dollar increase in the Facebook advertising budget.

   β_2 is the average change in sales for a one dollar increase in the Facebook advertising budget, assuming that the YouTube advertising budget is held constant.

   β_2 is the exact change in sales for a one dollar increase in the Facebook advertising budget, assuming that the YouTube advertising budget is held constant.

   β_2 is the average change in sales for a one thousand dollar increase in the Facebook advertising budget, assuming that the YouTube advertising budget is held constant.

4. Consider a dataset containing three variables, each measured in dollars: 4 / 4 points

   - Sales of a product (response, Y)
   - YouTube advertising budget (predictor, x_1)
   - Facebook advertising budget (predictor, x_2)

   Suppose that the relationship between the predictors and the response is linear, but that the predictors x_1 and x_2 are first scaled before modeled. That is:

   $$Y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \varepsilon$$

   where

   - $z_j = \frac{x_j - mean(x_j)}{sd(x_j)}, j = 1, 2;$
   - $mean(x_j)$ is the sample mean of the predictor x_j; and
   - $sd(x_j)$ is the sample standard deviation of the predictor x_j.

   What is the correct interpretation of β_0?

   β_0 is the mean/expected sales for the sample average Facebook and YouTube advertising budgets.

   β_0 is the mean/expected change in sales.

   β_0 is the mean/expected sales when all advertising budgets are zero dollars.

5. Consider a dataset containing three variables, each measured in dollars: 4 / 4 points

   - Sales of a product (response, Y)
   - YouTube advertising budget (predictor, x_1)
   - Facebook advertising budget (predictor, x_2)

   Which of the following most plausibly describes the statistical units in the dataset?

   The set of all companies that sell the product whose sales are quantified in Y.

   YouTube advertising budgets.

   Dollars.

   Facebook Advertising budgets.

   Individual companies that sell the product whose sales are quantified in Y.

6. Consider a dataset containing three variables, each measured in dollars: 4 / 4 points

- Sales of a product (response, Y)
- YouTube advertising budget (predictor, x_1)
- Facebook advertising budget (predictor, x_2)

Which of the following most plausibly describes the population of interest?

The set of all dollars spent on the product whose sales are quantified in Y.

The set of all companies that advertise in Facebook.

The set of all companies that advertise in YouTube.

The set of all companies that sell the product whose sales are quantified in Y.

Individual companies that sell the product whose sales are quantified in Y.

7. Which of the following are assumptions of the linear regression model? 4 / 4 points

The existence of a linear relationship between the response variable and the model parameters.

The existence of a linear relationship between the response variable and the predictor variables.

The existence of a linear relationship between the response and the random error term.

Distinct random error terms must be independent.

Each response measurement is assumed to have the same variance.

8. Consider a dataset containing three variables, each measured in dollars: 4 / 4 points

- Sales of a product (response, Y)
- YouTube advertising budget (predictor, x_1)
- Facebook advertising budget (predictor, x_2)

Suppose that the relationship between the predictors and the response is linear, but that the predictors x_1 and x_2 are first scaled before modeled. That is:

$$Y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \varepsilon$$

where

- $z_j = \frac{x_j - mean(x_j)}{sd(x_j)}, j = 1, 2;$
- $mean(x_j)$ is the sample mean of the predictor x_j; and
- $sd(x_j)$ is the sample standard deviation of the predictor x_j.

What is the correct interpretation of β_1?

β_1 is the average change in sales for a one standard deviation increase in the YouTube advertising budget, assuming that the Facebook advertising budget is held constant.

β_1 is the average change in sales for a one dollar increase in the YouTube advertising budget, assuming that the Facebook advertising budget is held constant.

β_1 is the average change in sales for a one standard deviation increase in the Facebook advertising budget, assuming that the YouTube advertising budget is held constant.

β_2 is the average change in the YouTube budget for a one standard deviation increase in sales, assuming that the YouTube advertising budget is held constant.