Assignment 1 ᴬ↓

Start Assignment

Due Tuesday by 5:29pm   **Points** 10   **Submitting** a file upload

Available after Jan 18 at 5:30pm

Assignment Details

Study Background and Data Attributes

Recently, researchers in Portugal attempted to create a biomarker-based prediction model
to identify women with breast cancer. They used data that would typically be collected in
routine consultations among women. Their work was reported in the following
manuscript:

• Patricio et al Using Resistin, glucose, age and BMI to predict the presence of breast
cancer. BMC Cancer 2018; 18:29.

The dataset is a .csv file posted within Module 1.

Attribute Information from the Dataset

Quantitative Attributes:

Age (years)

BMI (kg/m2)

Glucose (mg/dL)

Insulin (µU/mL)

HOMA-IR (Homeostatic model assessment: method to assess beta-cell function and insulin resistance;
results in an index with high values indicating body is using more insulin than normal to keep blood sugar
in balance)

Leptin (ng/mL) (hormone level)

Adiponectin (µg/mL) (hormone level)

Resistin (ng/mL) (hormone level)

MCP-1(pg/dL) (chemokine level)

Classification Labels: 1=Healthy controls 2=Breast Cancer Patients

Assignment Instructions

Using the study dataset, you will complete the following tasks. Create and then submit, using R
markdown, a document that displays the **programming code** and **output associated with each item**.
Follow the formatting instructions reviewed in class. You must provide a specific answer for all items; you
are encouraged to output formatted and specific answers (e.g. only outputting a single estimate or
generating a formatted table). Use complete sentences for any short-answer questions.

1. Construct a table providing summaries of the quantitative features of the dataset. Summaries should include the mean, median, minimum value, and maximum value. If you are unable to construct a formatted table within R, you can print raw output, but then comment the output to identify the answer that was requested.

2. Recode BMI into the WHO-defined categories below
Severely underweight - BMI less than 16.5kg/m^2
Underweight - BMI under 18.5 kg/m^2
Normal weight - BMI greater than or equal to 18.5 to 24.9 kg/m^2
Overweight – BMI greater than or equal to 25 to 29.9 kg/m^2
Obesity class I – BMI 30 to 34.9 kg/m^2
Obesity class II – BMI 35 to 39.9 kg/m^2
Obesity class III – BMI greater than or equal to 40 kg/m^2

3.Create a bar chart showing the proportion of breast cancer cases and controls within each BMI category

4. Construct a logistic regression model using breast cancer classification as the outcome and glucose, HOMA-IR, leptin, BMI (continuous) and age as the independent variables. Fill in the beta estimate and 95% confidence interval associated with a 1-unit change in HOMA-IR

5.Construct a linear regression model using insulin as the outcome and BMI (continuous), age, and glucose as the independent variables. Fill in the beta estimate and 95% confidence interval associated with a 1-unit change in age.