In the name of God

Quiz 2– Data science and HPC

- Write your codes in a notebook file so we can see codes and results together.
- If you want to write your codes in a jl file, take a couple of photos of your terminal and attach them in a zip file.

1) plot Y = 2*x + e, when your error, e ~ $N(0,1)$ and 0 < x < 100. (x ∈ N).

2) Construct a function that calculates the factorial for the input number (0! = 1, 1! = 1) and prints an error "your input variable is not valid" if the input number doesn't belong to the valid collection, for example, 2.5! or negative numbers.

3) "dataset.xlsx" is taken from a public survey on whether a person ever had an accident or not. Now, as a data scientist, we want to propose a model and measure its accuracy, assume we select logistic regression for that.

a. First randomly select 15% data as test set and set remaining as train set.
b. some records in Age column are null value you must fill it, one of the good choice is filling with average of remaining data.
c. After fitting model, test model and evaluating the model, use confusion matrix (Precision, recall) to evaluating your model.

4) Give an example that if learning rate α is chosen inappropriately, the gradient descent diverges.

5) Consider the conjugate gradient descent algorithm method (look it up in toward data science).

a) Let M be a symmetric, positive definite square matrix with n rows and columns. If the vectors $v_1, \ldots, v_n$ are M-conjugate, then these vectors are linearly independent and span the space R^n.
b) Explain why the iterations of the conjugate gradient descent method are finite.

Good luck