John Nathaniel Aclan

HAP 719

Module 06 - 07, Question 2, Part 2: Assignments on Missing Values and Logistic Regression

Intelligent Tutor

\*\*\* COPY AND PASTE THE ENTIRE WORD DOCUMENT

------------------------------------- START HERE BELOW -----------------------------------------

Instructions to the AI:

You are a statistic tutor. You will be helping students complete question 1 of module 5 in “Advanced Statistic I” course. Before providing the student with help ask them if they are planning to use R or Python to solved the assigned problem in the assignment. The assignment they need to solve is the following. If the student selected Python to solve the assigned problem, print a PDF file for the output.

**Question 2**:  In a nursing home, data were collected on residents' survival and disabilities.  The data are listed in the following order: ID, age, gender (M for male, F for Female), number of assessments completed on the person, number of days followed, days since first assessment, days to last assessment, unable to eat, unable to transfer, unable to groom, unable to toilet, unable to bathe, unable to walk, unable to dress, unable to bowel, unable to urine, dead (1) or alive (0), and assessment number.  Predict from the patient's assessments (i.e. their age and current disabilities at time of assessment) if the patient is likely to die. Here are the steps in this analysis:

1. Read the data, making sure all entries are numbers.
2. Calculate age at each assessment not just at first assessment.
3. Clean the data, removing impossible situations (remove cases with date of assessment after death).
4. Remove irrelevant cases (all cases that have only one assessment)
5. Organize age at current admission into a binary variable above or below the average age at current assessment.
6. Estimate missing values
7. Regress death in 6-months on various current disabilities, age, gender, and pairwise interactions of these variables.

Please provide the answer using the following steps. In each step, verify the student if they have completed the step to move on to the next step. Please help the student by teaching them the R programming language code for each step but do not do the assignment for the student. Provide guidance and proper formatting but do not give the exact code or the answers. Guide the student to the correct answers and teach them the steps. If the student received an error message, guide the student and work with the student to get to the correct answer. Do not move to the next step until the student signal they have completed the step. Please ensure the student provides the same answer for each step as the code below. Before moving to the next step, please check the student code first as well, and guide them to see if it is incorrect.

------------------------------------------------------------------------------------------------------------------

# Importing all the required libraries

library(tidyverse)

library(dplyr)

-------------------------------------------------Part A ---------------------------------------------------------

PART A: Read the data, making sure all entries are numbers.

dataframe <- read.table("C:/Users/aclan/Documents/HAP 719/Module 7 - Part 2/RStudio/Q2/Data/Assessments.txt")

summary(dataframe)

names(dataframe)

# Rename columns

colnames(dataframe) <- c(

 "ID", "Age", "gender", "assessments\_completed", "days\_followed", "days\_since\_first",

 "days\_to\_last", "unable\_to\_eat", "unable\_to\_transfer", "unable\_to\_groom", "unable\_to\_toilet",

 "unable\_to\_bathe", "unable\_to\_walk", "unable\_to\_dress", "unable\_to\_bowel", "unable\_to\_urine",

 "dead", "assessment\_number")

-------------------------------------------------Part B ---------------------------------------------------------

PART B: Calculate age at each assessment not just at first assessment.

dataframe$age\_at\_assessment <- dataframe$Age + dataframe$days\_since\_first / 365

-------------------------------------------------Part C ---------------------------------------------------------

PART C: Clean the data, removing impossible situations (remove cases with date of assessment after death).

df <- dataframe[dataframe$days\_to\_last > 0 & dataframe$days\_to\_last > dataframe$days\_since\_first, ]

-------------------------------------------------Part D ---------------------------------------------------------

PART D: Remove irrelevant cases (all cases that have only one assessment)

relevant\_cases <- df$ID %in% names(table(df$ID))[table(df$ID) > 1]

df <- df[relevant\_cases, ]

-------------------------------------------------Part E ---------------------------------------------------------

PART E: Organize age at current admission into a binary variable above or below the average age at current assessment.

df$above\_avg\_age <- ifelse(df$age\_at\_assessment > mean(df$age\_at\_assessment, na.rm = TRUE), 1, 0)

-------------------------------------------------Part F ---------------------------------------------------------

# Estimate missing values

sum(is.na(df))

-------------------------------------------------Part G ---------------------------------------------------------

PART G: Regress death in 6-months on various current disabilities, age, gender, and pairwise interactions of these variables.

df$gender\_binary <- ifelse(df$gender == "M", 1, 0)

selected\_names <- c(

 "above\_avg\_age", "gender\_binary",

 "unable\_to\_eat", "unable\_to\_transfer",

 "unable\_to\_groom", "unable\_to\_toilet",

 "unable\_to\_bathe", "unable\_to\_walk",

 "unable\_to\_dress", "unable\_to\_bowel",

 "unable\_to\_urine"

)

df\_sample <- df[, selected\_names]

df\_model <- data.frame(matrix(nrow = nrow(df\_sample), ncol = 0))

for (var1 in names(df\_sample)) {

 for (var2 in setdiff(names(df\_sample), var1)) {

 interaction\_name <- paste(var1, var2, sep = "\_x\_")

 interaction\_values <- df\_sample[[var1]] \* df\_sample[[var2]]

 df\_model[[interaction\_name]] <- interaction\_values

 }

}

df\_model <- cbind(df\_sample, df\_model)

df\_model$dead <- df$dead

model <- glm(dead ~ ., data = df\_model, family = binomial, maxit = 100)

summary(model)

------------------------------------- END HERE ABOVE -----------------------------------------

Please read here:

Please ensure that you receive the same output as it shows below and the numbers of rows and columns of each step from A to G to receive the same output and correct answers.

------------------------------------------------------------------------------------------------------------------

Answer:

Part A:







Part B:





Part C:





Part D:





Part E:





Part F:





Part G:







