Name:krutika

Course NO.: HAP 719

**MISSING VALUES**

|  |  |
| --- | --- |
| Question NO. | Correct Answer |
| 1 | Yes |
| 2 | Yes |

Regress progression in Infectious and Parasite body system on all other variables (except diabetes). In the attached data, the variables indicate incidence of diabetes (a binary variable) and progression of diseases in body systems. You can do the analysis first on 10% sample.

1. a) Remove all instances where progression in Infectious and Parasite body system is missing.  Report the number of cases that remain.

A close-up of a computer screen

Description automatically generated

1. b) Plot the progression in Infectious and Parasite body system. Is the data bimodal?  Is the data symmetric around the mean? Transform the data to improve QQplot

A collage of graphs and diagrams

Description automatically generated

* c) C) Include in your analysis all pairwise, and triplets of the independent variables.  Exclude any variable or interaction term that is always missing. What is the total number of independent variables included in the regression?

A computer screen shot of a computer code

Description automatically generated

d) d) Impute missing independent variables from other variables that are present.  Regress progression in Infectious and Parasite body system on independent variables and report the percent of variation explained

A screenshot of a computer

Description automatically generated

e) Assume that missing independent variables indicate the patient does not have any disease in the body system (i.e., 0 score).   Regress progression in Infectious and Parasite body system on independent variables and report the percent of variation explained.

A screenshot of a math problem

Description automatically generated

f) Assume the value of the missing independent variables can be replaced by the average value of the independent variable. Regress progression in Infectious and Parasite body system on independent variables and report the percent of variation explained.

A screenshot of a computer

Description automatically generated

* QUESTION:2

 Consider the regression of progression in Infectious and Parasite body system on all other variables (except diabetes). In the attached data, the variables indicate incidence of diabetes (a binary variable) and progression of diseases in body systems. You can do the analysis first on 10% sample before you do it on the entire data that may take several hours.

A screenshot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated