SAS macro

What can macro help?

1. change the model parameters and input with just 1 statement. ( using macro variable)

2. reuse the modular code.

Whats data mining?

mine knowledge from large amount of data

Step for knowledge discovery:

1. data cleaning

tackle:

noisy(errors and outliers) smoothing, outlier detection

missing

inconsistent

2. data integration

combine data from multiple sources

3. data reduction

attain reduced representation of dataset in smaller volume

method:

1. attribute subset selection: irrelevant, weakly related attributes are removed, such as stepwise forward/backward

2. dimensionality reduction: encoding mechanisms are used to reduce the data set size

e.g, data compression/ data structure

3. data cube aggregation

4. data transformation

smoothing/aggregation/generalization/normalization

5. data mining

characterization: summarize data of a class in general forms

discrimination: compare the target class with other classes , just like compare software which has increasing sale with that has decreasing sale

frequent pattern: frequent itemsets, market basket analysis

grequent subsequences, event A follows B

association

classification/prediction

cluster analysis