Q1) In your own words define multivariate analysis?

Multivariate analysis is a superset of all statistical techniques and methods that help us simultaneously analyze data of multiple dimensions/ data sets/ variables to find the relationships between them and we can create visualizations to understand the various variables of data sets to get a better understanding on a whole another level.

Multivariate – Working with multiple variable's.

They can work with two or more variable at a time.

Analysis- is done to find the relationship between these variables.

Q2) Name the most important factors contributing to the increased application of techniques for multivariate data analysis in the last decade.

The most important factors that contributed to increased application are:

- Lot of data accumulated over the data warehouses from databases, operational systems.
- Data mining for improved decision making
- High speed computing performance
- Technology advancements help us apply various new multivariate functions over large data sets and process them with ease.
- The improved computing capacity to conceptualize information from such huge data sets and visualizing them. Even the software and hardware components of computers have improved significantly to help us perform such tasks.
- The development of computer based statistical methods to analyze, plot and visualize data.

Q3) List and describe the multivariate data analysis techniques described in this chapter. Cite examples for which each technique is appropriate.

Basically, they are mainly classified into two types of methods.

1. Dependent - analyze dependent and independent variables at the same time.
-
2. Independent - analyze dependent and independent variables separately

Q4) Explain why and how the various multivariate methods can be viewed as a family of techniques.

The various multivariate methods can be viewed as a family of techniques as they all basically work to build complex relationships amongst the various variables / components / datasets. The various multivariate statistical methods accompany each other by having relative functions and combinations of the input data set and output projection needs and requirements. So, these various capabilities can be brought together to work closely as a family and inter related important techniques and functions. Thus, these can be used to solve complex data having more than 2 variables and large data sets.

Basically, all multivariate techniques are divided based on 3 major factors:

1] Dependent or Independent

2] Dependency over number of variables for dependent methods

3] Metric or non-metric for Interdependence methods.

Based on these factors we choose the best fit multivariate method.

At times when we have a condition that fits more than one multivariate method then both offer the same output thus we can call them a family of techniques.

For Example: if we have a Several dependent variables and classifies as metric then we can use MANOVA or Canonical.

Q5) why is knowledge of measurement scales important to an understanding of multivariate data analysis?

Knowledge of measurement of scales is extremely important to get a better understanding of multivariate data analysis because application of wrong statistical methods could lead to inaccurate or wrong results. A proper understanding of the data and the variable relations and requirements should be understood properly to apply the required method to derive the relation that the problem requires and so that the data can be analyzed appropriately which can further help us derive important relations from the given data .

Q6) what are differences between statistical and practical significance? Is one a prerequisite for the other?

Statistical significance is a theoretical calculation done over the data set to determine the requirements and conditions and find the best fit for it.

- Type I error, or α, is the probability of rejecting the null hypothesis when it is true.
- Type II error, or β, is the probability of failing to reject the null hypothesis when it is false.
- Power, or 1-β, is the probability of rejecting the null hypothesis when it is false.

Practical significance usually is done when we have large sample sizes and the tiniest effect could seem to be statistically significant, so there we need to practically check for the significance.

Statistical significance should be the prerequisite for practical significance

For Ex: There could be an upper bound and lower bound limit test for weight of a toothpaste in the production line. Practically getting 1-2 grams over or below the specified limit is fine but according to the statistical significance it is wrong it would be rejected.

Q7) What are the implications of low statistical power? How can the power be improved if it is deemed too low?

If the statistical power is low then it will be difficult for the researcher or data analyst to analyze or find the important relations even though they exist. If the power is deemed too low then they can improve the power by decreasing the alpha levels or also by increasing the sample size.

Q8) Detail the model building approach to multivariate analysis, focusing on the major issues at each step.

The model building approach to multivariate analysis involves the following phases:

Stage 1: Define the Research Problem, Objectives, and Multivariate Technique to be Used

- This phase involves the creation of conceptual and objective model and then create the whole specifications about them. Further find the type of relationship i.e. whether it's an

dependent relationship or an independent relationship. Finally choosing the most appropriate multivariate technique.

Stage 2: Develop the Analysis Plan

Now we need to think about the implementation issues

Stage 3: Evaluate the Assumptions Underlying the Multivariate Technique

Stage 4: Estimate the Multivariate Model and Assess Overall Model Fit

Stage 5: Interpret the Variate

Stage 6: Validate the Multivariate Model

A Decision Flowchart: This is divided into 2 sections:

The first part involves stages 1-3, which deals with issues that occur while preparing for actual model estimation.

The second part involves stages 4-6, this deals with issues related to model estimation, interpretation and validation.

References :

Textbook : Multivariate Data Analysis

Book by Barry J. Babin, Hair, Rolph E Anderson, and William C. Black
Powerpoint : Lecture 1