

STAT 467 Multivariate Data Analysis

Project Guideline

As part of the requirements for STAT 467 Multivariate Analysis, you are expected to complete a course project. The purpose of this project is to apply the multivariate methods covered in the course to a real dataset, interpret the results, and present your findings in a clear and professional manner.

Groups:

Form your project teams with **3 students** (although working in pairs (2 students) or individually is also possible if necessary). Please choose one member of your team as the "**corresponding member**." This person will be responsible for communication between the group and the instructors. All other parts of the project must be completed jointly, with equal contribution from all team members. The corresponding member should send me an email (bkoca@metu.edu.tr) that includes the full names (name and surname) of all group members while specifying the corresponding person and information about the dataset you plan to use.

Deadline for informing: 21.11.2025, 17:00.

Data Sets:

Your data should have at least ten (10) continuous variables, and some categorical and discrete variables. The data should be big enough to apply multivariate techniques. If the data set is very big, you can use a randomly chosen subset of it. You can work with any data that interests you in accordance with these terms. If you wish, you can use the following resources to find a data set:

- The UC Irvine Machine Learning Repository
- The World Bank Microdata Library
- UNHCR Microdata Library
- FAO Microdata Catalogue
- Kaggle
- Google Dataset Search
- Kegg

Each group must select a new dataset for this semester. Datasets that were used in previous years are **not** allowed. Additionally, each group must work with a different dataset, so **NO** two groups may use the same data. When you submit your group members' names and information about your chosen dataset to the assistant, you will be notified if the dataset has been used in previous years or chosen by other teams in the current semester.

Methods to Be Applied in the Project:

1. Exploratory Data Analysis
2. Inferences About a Mean Vector
3. Comparisons of Several Multivariate Means
4. Principal Components Analysis and Principal Components Regression
5. Factor Analysis and Factor Rotation
6. Discrimination and Classification
7. Clustering
8. Canonical Correlation Analysis (Bonus)

Caution: Please make sure that the analyses should be conducted from your side. Do not copy and paste the existing analyses from known literature data. This will be checked through TURNITIN program. If plagiarism is detected in the project, all members of the group will receive a grade of zero.

Notes:

- (i) There is NO written report requirement for this project. Instead, each group will give an oral presentation. Your presentation should include all steps of your analysis, from data preparation to the application of multivariate methods and interpretation of results. In addition, each group must submit the presentation file, the updated version of dataset and codes to ODTUClass. Please don't send any of these files via an e-mail.
- (ii) Only the corresponding member should upload the required files to ODTUClass; the other group members do not need to upload anything. Please change the name of presentation file as "Stat 467 Project Presentation Group X", instead of "X" write your group number which will be assigned.
- (iii) Add all the references you used for this project. Do NOT forget to add your data sources, as well.

Oral Presentation:

This presentation should contain

- Exploratory Data Analysis
 - ❖ Data Description
 - ❖ Visualization Tools
 - ❖ Problem/Main Research Questions
- What You Have Done (Implementations) (Please avoid including unnecessary or irrelevant information in your presentation. Focus only on the key steps of your analysis.)
- Main Conclusions

Notes:

- (i) The group will give a 15-minute oral presentation together, covering all slides. Please include names and surnames of all group members in the first page of presentation.
- (ii) The oral presentations will be held on the last week of December (and/or the first week of January).

Evaluation Criteria for Presentation

The presentations will be evaluated according to the following criteria:

Category	Scoring Criteria	Total Points
Organization (20 points)	The introduction gave a clear idea of the topic and established a framework for the rest of the presentation.	5
	Information was presented in a logical sequence which audience can follow.	5
	The materials (Visuals/Graphics) were a clear and structured manner.	5
	The conclusion did a good job of summarizing the presentation.	5
Content (60 points)	Technical terms were well-defined.	15
	Enough essential information was given to effectively evaluate the topic.	15
	The content was free from irrelevant information.	15
	Material included was relevant to the main points of the presentation.	15
Presentation (20 points)	Delivery was poised, controlled, and smooth.	5
	The presenter was well organized and prepared.	5
	Information was well communicated.	5
	Visuals were well prepared, informative, effective, and not distracting.	5
	Length of presentation was within the assigned time limits.	5
TOTAL SCORE		100

Important Dates:

The group members, specifying the corresponding person, and a data set (with website URL) for the project should be determined and reported by November 21st, 17:00 PM via e-mail to the research assistant (bkoca@metu.edu.tr).

The project presentation (1), R/Python code files (2) and the updated version of the data set (3) should be uploaded to ODTUClass by December 29th, 11:59 PM. **NO late submissions will be accepted and evaluated.**

If you have any questions regarding the project, please feel free to contact the instructor team (vpurutcu@metu.edu.tr & bkoca@metu.edu.tr).