Department of …Statistics & Informatics….

College of Administration and Economics

University of Salahaddin-Erbil

Subject: Multivariate Statistical Analysis

Course Book: 4th Year (Senior)

Lecturer's name: Dr. Delshad Shaker Botani (PhD)

Academic Year: 2021 – 2022

Course Book

1. Course name	Multivariate Statistical Analysis
2. Lecturer in charge	Dr. Delshad Shaker Ismael Botani
3. Department/ College	Statistics & Informatics /Adm. & Eco.
4. Contact	e-mail: **delshd.botani@su.edu.krd** Tel: (optional)
5. Time (in hours) per week	Theory: 3 Practical: 0
6. Office hours	Tuesday from 10:30 AM to 4:30 PM Wednesday from 8:30 AM to 1:30 PM
7. Course code	
8. Teacher's academic profile	From 1996 until 2000, I worked as an assistant lecturer in Statistics department – Salahaddin University. In 2002, I had MSc. in Statistics from the same university. Then, from 2004 until the end of 2005 I worked in the Mathematics department – Science College teaching some statistical subjects such as Statistics, Probability, Math. Statistics, SPSS, with supervision of students' graduation projects. At the same time, I was the head of Computer unit at Science College. After that I accepted in the PhD scholarship in Russia and commenced the study in Moscow State University for Economics, Statistics, and Informatics (MESI) at the end of 2005. In 2010, I had my PhD in Statistics from the MESI university. Since 2010 I am working as a lecturer in Statistics department – Salahaddin University teaching statistical subjects and supervising BSc. and MSc. students. Also, I was a lecturer in the College of Police in Erbil city teaching computer science subject. I am the owner of one published book written in Kurdish language named "Amarzani" which means Statistics. I have more than 12 scientific researches in Statistics field. From 2017, my scientific level changes to assistant Professor. During my presence in this field, I have taught the following subject: Statistics, Visual Basic, SPSS, Probability, Math. Statistics, Operations Researches, Research Methodology, Surveys, Multivariate Statistics. I am a member in the following organizations: ✓ Head of STATISTICS Organization for Society Support (SOSS) in KRG (From 2021 – Till now). ✓ ESOMAR (ESOMAR is the essential organization for encouraging, advancing and elevating market research worldwide (From 2012 – Till now); https://www.esomar.org/community/our-community/member-search/profile681531_Delshad-Botani.php ✓ KURDISTAN ECONOMIST SYNDICATE in KRG (From 2010 – Till now): Conducting feasibility studies of many sectors, such as: industry, real estate, agriculture, Banking, Health care, Education, Oil and gas producers, etc. More than 100 feasibility studies conducted from 2011 till now.
9. Keywords	Multivariate, Quadratic form, Multiple Correlation Coefficient, Hotelling T^2, MANOVA, Factor analysis, cluster analysis, Multiple Regression Analysis, Discriminant Analysis

10. Course overview:

Most of the observable phenomena in the empirical sciences are of a multivariate nature. In the past 20 years, with the computer application technology and the urgent need for research and production, multivariate statistical analysis techniques are widely used in geology, meteorology, hydrology, medicine, industry, agriculture and economic and many other fields, has become to solve practical problems in effective way.

Multivariate statistics refer to an assortment of statistical methods that have been developed to treat situations in which multiple variables are involved. Any analysis of more than two variables can loosely be considered a multivariate statistical analysis.

So, how can we introduce our students to big data sets and basic techniques for multivariate data analysis when those students have little background in big data or multivariable manipulations? This course will try to answer this question and teach students the most important methods of multivariate data analyses, such as multivariate normal distribution, multiple correlation coefficient, MANOVA, Factor Analysis, etc. Computer literacy is essential, as we make extensive use of the computer using statistical software SPSS.

11. Course objective:

The following are the most important objectives of this course:

A. Introduce the language of multivariate data analysis.
B. To gain a thorough understanding of the details of various multivariate techniques, their purposes, their assumptions, their limitations, and so on. Many of these techniques are related; yet they differ in some essential ways. We emphasize these similarities and differences.
C. To be able to select one or more appropriate techniques for a given multivariate data set. Recognizing the essential nature of a multivariate data set is the first step in a meaningful analysis.
D. To be able to interpret the results of a computer analysis of a multivariate data set using SPSS Package (v.25).

12. Student's obligation

Students are expected to:

- Follow university policies when attending class and lab, and taking sudden quizzes and exams.
- Student should be proud of the work that he/she do in this class. Do not allow someone else to copy your homework and do not provide answers to quizzes or tests. If this does occur, credit will be lost and a referral will be written.

13. Forms of teaching

We will focus on some forms of teaching such as classical teaching with whiteboard, PowerPoint presentations for the head titles, definitions and summary of conclusions, classification of materials and any other illustrations, solving the examples by sharing the students to get them will understand, and students should participate as much as possible in lecture's discussions.

14. Assessment scheme

The students are obliged to perform at least two closed book exams during the academic year. The exam has approximately 30%, besides homework, quizzes, and classroom activities about – 10%. The other 60% will be reserved for the final exam. Therefore, the final grade will be based upon the following criteria:

Homework and interactive activities: 10%
Theoretical Exam : 30%
Final Exam : 60%

15. Student learning outcome:

After the completion of this course in this academic year, the students will be able to do the following:

- Appreciate the range of multivariate techniques available;
- Summarize and interpret multivariate data;
- Understanding of the link between multivariate techniques and corresponding univariate techniques;
- Using multivariate techniques appropriately, undertake multivariate hypothesis tests, and draw appropriate conclusions.

16. Course Reading List and References:

1) Daniel J. Denis (2019). *SPSS data analysis for univariate, bivariate, and multivariate statistics*. NJ: Wiley, 2019
2) Keenan A. Pituch & James P. Stevens (2016). *Applied Multivariate Statistics for the Social Sciences: Analyses with SAS and IBM's SPSS (6th ed.).* Routledge
3) ALVIN C. RENCHER, William F. Christensen (2012). *Methods of Multivariate Analysis* (3rd ed.). John Wiley & Sons.
4) Johnson, Richard A.; Wichern, Dean W. (2007). *Applied Multivariate Statistical Analysis (Sixth ed.).* Prentice Hall.
5) Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis (3rd ed).* Wiley, New York.
6) Siotani, M., Hayakawa, T., and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis.* Columbus, Ohio: American Sciences.
7) Eaton, M. L. (1983). *Multivariate Statistics.* Wiley, New York.
8) Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, London.
9) Bock, R. D. (1975). *Multivariate Statistical Methods in Behavioural Research.* New York: McGraw-Hill.

17. The Topics:	**Lecturer's name**
1) Introduction and overview of the course. Some principles of multivariate statistical modelling. 2) Matrices and Linear Algebra. 3) Variance-Covariance Matrix (Random Vectors, means, variance and Covariance Standard Scores matrix) 4) Correlation matrix (Simple Linear Correlation, Partial Correlation, Multiple Correlation) 5) Test of Correlation (Test for Population Correlation Coefficient equals Zero, Test for Population Correlation Coefficient equals a specified value, Test for Population Partial Correlation Coefficient, Test for Equality of two Population Correlation Coefficients, Test for Multiple Correlation Coefficient) 6) Eigen values and Eigen Vector 7) Characteristic equation (Properties of Eigen values and Eigen Vector) 8) Quadratic Form (Quadratic Form in n multivariate, Classification of matrix, Some Properties of Quadratic Form for Symmetric matrix) 9) Multivariate Normal distribution (Generalized Univariate Normal distribution, Bivariate normal distribution, Multivariate Normal distribution, Properties of Multivariate Normal random variables) 10) Test for population Mean Vectors (Test for population Mean Vectors (Covariance Matrix Unknown), Test for Equality of Population Means vectors (Covariance Matrices are Equal and known), Test for Equality of Population Means vectors (Covariance Matrices are Equal and Unknown), Test for Equality of Population Means vectors (Covariance	Dr. Delshad Botani Three hours a week

Matrices are Unequal and Unknown) 11) Hypothesis Tests for Covariance matrix (Hypothesis Test for Single Covariance matrix, Tests Comparing Covariance Matrices) 12) Principle Components Analysis 13) Discriminate Analysis (Description of Group Separation, The Discriminate Function for Two Groups, Relationship between Two-Group Discriminate Analysis and Multiple Regression, Discriminate Analysis for Several Groups, Discriminate Functions, A Measure of Association for Discriminate Functions) 14) Multivariate Regression (Least Squares Estimation in the Fixed-x Model, The Model Corrected for Means, Subset Selection)	
18. Practical Topics (If there is any)	

19. Examinations:

I. **Compositional:** This type of exams usually starts with define, Explain How, What are the reasons for…?, Why…?, How….? With their typical answers. The following questions are some examples of this kind of examinations:

Define Multivariate Analysis.
Describe multivariate statistical analysis and give a real daily example of it.
What is principle component analysis?
What is the propose of Discriminant analysis?

II. **Fill the blanks:** This type of question or phrase with one or more words replaced with a blank line, giving the student the chance to add the missing word(s), such as the following questions:

- ✓ The most popular rotation strategy in factor analysis is _VARIMAX or orthogonal rotation_.
- ✓ _Partial correlation__ is a correlational technique that allows you to evaluate the relationship between two variables with the effects of a third removed from both of them.
- ✓ The number of possible discriminant functions in a discriminant analysis is limited to _number of levels of the dependent variable minus 1_ or to number of predictors __, whichever is less.
- ✓ In the formula for the Pearson correlation coefficient, the calculation of variability can be found in _the numerator_

III. **True or false type of exams:** In this type of exam, a short sentence about a specific subject will be provided, and then students will comment on the trueness or falseness of this particular sentence, such as the following questions:

- ➢ _T__ In multivariate statistical analysis, the word "multivariate" indicates that two or more dependent variables are simultaneously used in a single analysis.
- ➢ _F__ Correlation for multiple regression predictions always infers causation. **(Not infers causation, it is not a must)**
- ➢ _F__ The higher the multicollinearity, the smaller the standard error for the regression coefficients.**(the higher the multicollinearity, the larger the standard errors)**
- ➢ _T__ Data errors and observations which represent extreme magnitudes on variables are both causes of outliers.

IV. **Multiple choices:** In this type of exam there will be a number of phrases next or below a statement, students will match the correct phrase, such as the following questions:

- ❖ __ used when a numerical predictor has a curvilinear relationship with the response.
  a. Simple regression *b. Multiple regression* *c. Quadratic regression* *d. all of them* *e. None of them*
- ❖ ____ used to check the assumptions of the regression model.
  a. Stepwise regression *b. R^2 Adjusted* *c. Correlation* *d. all of them* *e. None of them*
- ❖ ____ worst kind of outlier, can totally reverse the direction of association between x and y.
  a. Residual plots *b. independent variable* *c. dependent variable* *d. bar chart* *e. None of them*
- ❖ __ problem that can occur when the information provided by several predictors overlaps.
  a. Cause and effect *b. Logistic regression* *c. Factor analysis* *d. Error* *e. None of them*

V. **Proofs and solutions:** Any proofs or solutions for practical questions may be provided in the exam and students should solve those questions or proof those theorems that are given at daily lectures (It is possible to solve this type of questions in any way), such as the following questions:

If $B = \begin{bmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{bmatrix}$, Show that B is idempotent matrix.

Solution:

$B.B = B$

$$B.B = \begin{bmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{bmatrix} . \begin{bmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{bmatrix}$$

$$B.B = \begin{bmatrix} \frac{16}{25} + \frac{4}{25} & -\frac{8}{25} - \frac{2}{25} \\ -\frac{8}{25} - \frac{2}{25} & \frac{4}{25} + \frac{1}{25} \end{bmatrix}$$

$$B.B = \begin{bmatrix} 20/25 & -10/25 \\ -10/25 & 5/25 \end{bmatrix}$$

$$B.B = B = \begin{bmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{bmatrix}$$

VI. **Home Examinations:** This type of exams will be conducted at home through sending an email to the students and they must answer the questions in a determined period or time, then all students must send their answers through the same email, such as the following questions:

Theorem: For every vector $\underline{\alpha}$ if $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$are independent. Let X_i be any Component of $\underline{X}^{(1)}$, Show That for all linear Combination $\underline{\alpha}\ \underline{X}^{(2)}$ Which is Minimize the Variance $\left(X_i - \underline{\alpha X}^{(2)}\right)$.

1. **Proof:** $V\left(X_i - \underline{\alpha X}^{(2)}\right) = V(X_i) + V\left(\underline{\alpha X}^{(2)}\right) - 2Cov\left(X_i, \underline{\alpha X}^{(2)}\right)$

   $= V(X_i) + \underline{\alpha}' V\left(\underline{X}^{(2)}\right)\underline{\alpha} - 2\underline{\alpha} Cov\left(X_i, \underline{X}^{(2)}\right)$

   $\sum_i = \begin{bmatrix} \sigma_{ii} & \sigma_i \\ \sigma_i' & \sum_{22} \end{bmatrix}$

   $= \sigma_{ii} + \underline{\alpha}' \sum_{22} \underline{\alpha} - 2\underline{\alpha}\sigma_i$

   $\frac{\partial\left(V(X_i - \underline{\alpha X}^{(2)})\right)}{\partial \underline{\alpha}} = 0 + 2\underline{\alpha}\sum_{22} - 2\sigma_i = 0$

   $\underline{\alpha}\sum_{22} - \sigma_i = 0 \qquad) * \sum_{22}{}^{-1}$

   $\underline{\alpha}\sum_{22}\sum_{22}{}^{-1} - \sigma_i \sum_{22}{}^{-1} = 0$

   $\underline{\alpha} = \sigma_i \sum_{22}{}^{-1}$

   $\underline{\beta} = \underline{\alpha} = \sigma_i \sum_{22}{}^{-1}$ Prove

20. Extra notes:

If any student cannot make it to an in class exam due to a documentable reason, please let me know as soon as possible. Makeup will not be allowed for home works. However, I will double count students future graded assignments in the cases of excused absences.

If you have any suggestions or concerns, either positive or negative, about this class, please do not hesitate to see me during my office hours or make an appointment. It is my hope that I will be able to resolve the issue.

21. Peer review

Dr. Saman Husein Mahmoud

Dr. Rizgar Magdid Ahmed