Syllabus: EMSE 6765

Course and Contact Information:

Course: EMSE 6765.10 – Data Analysis for Scientists and Engineers

Semester: Fall 2016 # of Credit Hours: 3.0

Meeting Time: Tuesday from 6:10PM to 8:40PM

Location: Tompkins Hall 406 – DSS Lab (22 Computers available)

Instructor:

Name: J. René van Dorp, Professor

Campus Address: 800 22nd Street, Office 2800, Washington DC 20052

Phone: 202-994-6638

E-mail: dorpjr@gmail.com

Office hours: Wednesday 2:00PM to 4:00PM

Course Description:

Topics that will be discussed include estimation, confidence intervals, hypothesis testing and goodness-of-fit testing. These methods perform statistical inference in a single dimension (also known as univariate data analysis). Discussions of multivariate data analysis utilize matrices and vectors. Rules of matrix-vector algebra will be reviewed and some intuitive geographical interpretations of these operations will be provided. Multivariate data analysis will be introduced via the classical Hotelling T^2 hypothesis test, which is a natural extension of the univariate T test. Next, regression analysis (in matrix-vector format), principal component analysis and analysis of variance are introduced. Their application will be facilitated by the use of the MINITAB software program. Discussion of these multivariate techniques will concentrate on intuition and conceptual understanding, not a rigorous mathematical derivation of their methodologies.

Total of 112.5 Student Engagement Hours are divided over: 2.5 hours of class instruction + midterm exam and a final report project over 15 weeks. Homework, reading assignments and preparation of final report is estimated at 5 hours per week over 13 weeks of class preparation. Studying/preparing for the midterm exam is estimated at 10 hours. **Estimated student engagement hours totals to 112.5.**

Prerequisite Requirement:

ApSc 3115: Engineering Analysis III (or any other undergraduate Applied Statistics course from a physical or natural sciences program). http://www.seas.gwu.edu/~dorpjr/APSC3115/Intro.html

Required Text: Electronic Lectures notes available at:
https://www.seas.gwu.edu/~dorpjr/EMSE271/Coursefiles.html

Required Software:

MS EXCEL – Available in Tompkins 406.
MINITAB – Available in Tompkins 406. Six months or twelve months rental of the MINITAB Software is available for students at a discounted rate at:
http://www.onthehub.com/minitab/

Recommended Text: Neither of the textbooks below is required. However, electronic lecture notes used throughout this course are developed from these course texts. Reading accompanying chapter from these texts may further enhance understanding.

"A Modern Introduction to Probability and Statistics, Understanding Why and How" by F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä and L.E. Meester, Springer-Verlag, 2005.

"Analyzing Multivariate Data" by Lattin, Carroll and Green.

Learning Outcomes:

As a result of completing this course, students will be able to:

1. Perform univariate statistical inference techniques involving confidence intervals, hypothesis test, distribution fitting and goodness-of-fit testing. Students will learn to perform these inference techniques in MS EXCEL.
2. Perform multivariate statistical inference techniques involving estimation of the sample mean vector, the sample variance covariance matrix and use these to perform the Hotelling

T^2 hypothesis test on a single multivariate sample and two multivariate samples. Students will learn to perform these inference techniques in MS EXCEL.

3. Perform regression analysis involving multiple explanatory variables using matrix algebra in MS EXCEL. Student will learn to perform and interpret regression analysis results using the software MINITAB
4. Perform principal component analysis by evaluating the eigenvectors and eigenvalues of the sample correlations matrix and use these to evaluate principle component loadings and variances. Introduces scree plots and loadings plots to facilitate principle component interpretations. The eigenanalysis of the sample correlation matrix is performed using the software MINITAB.
5. Perform One-Way, Two-Way and 2^K ANOVA to determine contributing factors in an engineering data analysis. One-Way and Two-Way ANOVA are performed using MINITAB. 2^K ANOVA is introduced using a MS EXCEL template.

Method of Instruction:

One hour and 15 minutes lecture including homework discussion, followed by a 15 minute break and a one hour lecture. Microsoft Excel and Minitab are used to perform statistical analysis during the class sessions and the homework. During class sessions the only software programs that should be open on your desktop are either Adobe Acrobat (for viewing the notes) or Microsoft Excel or MINITAB for statistical analysis. During the class sessions (except for the break of course) a student is not to check his e-mail, the internet and should not engage in instant messaging sessions. Basically, your attention should be directed towards the class material.

Grading:

20% - Class Attendance

40% - Midterm Exam (In-Class)

40% - Multivariate Data Analyses + Final Report

Homework:

Homework will have to be completed prior to the next class for discussion. The homework will not be graded, but is assigned to enhance your understanding of course materials and prepare yourself for the exams. Electronic solutions of the homework will be provided.

Reading Assignments:

Lecture notes and recommended chapters for reading will be assigned prior to class as indicated in the outline.

Midterm Exam and Final Report:

Students will complete an **in-class simulation Midterm Exam using Microsoft Excel + MINITAB** (using a lab computer or the student's laptop). Theoretical questions will be answered in an exam booklet. The MS EXCEL file, the MINITAB file and the exam booklet will be part of the grading of the midterm exam. Multivariate datasets will be provided to the students for analysis. Students will be required to perform multivariate data analyses using those datasets and write a final report detailing their analysis steps, final analysis results and analysis conclusions. Students are required to submit the electronic files associated with the final report as well as a hard copy of the final report that will be graded.

Lecture Notes:

Electronic copies of the lecture notes can be downloaded from my Faculty web-page at:
https://www.seas.gwu.edu/~dorpjr/EMSE271/Coursefiles.html

Class Schedule: Subject to change, please check the schedule regularly

	Session	Date	Day of Week	Reading Assignments	Topics	Homework Assigned
Part 1: Statistical Review	1	30-Aug	Tuesday	15 , 16	Exploratory data analysis: Graphical + Numerical Summaries	15.2, 15.10, 16.7, 16.13
	1	30-Aug	Tuesday	17	Basic Statistical Models	Homework Set 1
	2	6-Sep	Tuesday	19 - 20	Unbiased estimators, Efficiency and MSE	19.2, 19.7, 20.6, 20.9
	2	6-Sep	Tuesday	23	Confidence intervals for the mean: Essentials	Homework Set 2
	3	13-Sep	Tuesday	LN S3	Estimator distribution, Confidence Intervals for mean and Variance	
	3	13-Sep	Tuesday	LN S3	Hypothesis Testing, MLE, MOM	Homework Set 3
	4	20-Sep	Tuesday	LN S4	Goodness-of-Fit, Credibility Intervals	
	4	20-Sep	Tuesday	LN S4	Two Sample Hypothesis Testing, Joint Normal Distribution	Homework Set 4 Part I
Part 2: Hotelling	5	27-Sep	Tuesday	LN S5	Vectors and Matrices, Matrix Algebra, Linear Combinations,	
	5	27-Sep	Tuesday	LN S5	Coordinate Systems, Geometric Interpretation	Homework Set 4 Part II
	6	4-Oct	Tuesday	LN S6	Joint Normal Distribution, Multivariate Point Estimation	Homework Set 5 Part I & II
	6	4-Oct	Tuesday	LN S6	Generalized Variance, Hotelling's T^2 Test	Practice Exam
REVIEW + MIDTERM EXAM		11-Oct	Tuesday		No Class Professor's Travel	
		11-Oct	Tuesday		No Class Professor's Travel	
	7	18-Oct	Tuesday		Discuss Solution HW Set 5 I and II	
	7	18-Oct	Tuesday		Discuss Solution Practice Exam	
		25-Oct	Tuesday		Fall Break	
		25-Oct	Tyesday		Fall Break	
	8	1-Nov	Tuesday		EXAM - PART 1 and PART 2	
	8	1-Nov	Tuesday		EXAM - PART 1 and PART 2	
Part 3: Regression	9	8-Nov	Tuesday		Discuss Solution Midterm Exam	
	9	8-Nov	Tuesday	LN S9	Simple Linear Regression, Model Testing, Parameter Inference	
	10	15-Nov	Tuesday	LN S10	Multiple Regression, Residual Diagnostics, Outlier Detection	Homework Set 6
	10	15-Nov	Tuesday	LN S10	Comparing Imbedded Models, Forecasting	
Part 4: ANOVA	11	22-Nov	Tuesday	LN S13	One-Way Analysis of Variance (ANOVA)	Homework Set 9
	11	22-Nov	Tuesday			
	12	29-Nov	Tuesday	LN S14	Two-Way ANOVA, 2K ANOVA	
	12	29-Nov	Tuesday		Make-Up Class	
Part 5: PCA	13	6-Dec	Tuesday	LN S11	Principal Component Analysis (PCA), Introduction, How it works	Homework Set 7
	13	6-Dec	Tuesday			
	14	13-Dec	Tuesday	LN S12	Principal Component Analysis (PCA) Case Study	Homework Set 8
	14	13-Dec	Tuesday			
	15	20-Dec	Tuesday		FINAL PROJECT REPORT PART 3, 4 and 5 DUE	
	15	20-Dec	Tuesday		FINAL PROJECT REPORT PART 3, 4 and 5 DUE	

University Policy on Religious Holidays:

1. Students should notify faculty during the first week of the semester of their intention to be absent from class on their day(s) of religious observance.
2. Faculty should extend to these students the courtesy of absence without penalty on such occasions, including permission to make up examinations.
3. Faculty who intend to observe a religious holiday should arrange at the beginning of the semester to reschedule missed classes or to make other provisions for their course-related activities

Support for Students Outside the Classroom:

Disability Support Services (DSS)

Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services office at 202-994-8250 in the Rome Hall, Suite 102, to establish eligibility and to coordinate reasonable accommodations. For additional information please refer to: gwired.gwu.edu/dss/

Mental Health Services 202-994-5300

The University's Mental Health Services offers 24/7 assistance and referral to address students' personal, social, career, and study skills problems. Services for students include: crisis and emergency mental health consultations confidential assessment, counseling services (individual and small group), and referrals. counselingcenter.gwu.edu/

Academic Integrity Code

Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information. For the remainder of the code, see:

studentconduct.gwu.edu/code-academic-integrity