Stat 419 Introduction to Multivariate Statistics (Spring 18)
Washington State University

Instructor: Abhishek Kaul
Office: Neill Hall Room 217
Phone: (509) 335-5279
Email: akaul@math.wsu.edu
Class: Monday, Wednesday and Friday from 10:10 – 11:00am
Office Hours: Monday and Wednesday from 11:00am – 12:00pm, or by appointment
Prerequisite: Math 220 and one of STAT 360, 370, 412, 423 or 430.
Course Text: No required text, but recommended books given below.

Course Overview

This course will serve as an introduction to multivariate statistics. As more and more data are collected in multiple dimensions it is important that students understand the nuances and specialties related to data collected in multiple dimensions. The course will build on student’s basic knowledge of mathematics and statistics to introduce topics such as the multivariate normal distribution, principal component analysis, discriminant analysis. Implementation of the methods learned shall also be discussed in the statistical software R. Material learned in this course is useful for anyone entering fields such as data analytics, operations research, biological sciences and engineering. The skills learned in this course would be of particular interest to industries specializing in large multivariate data sets.

Learning Outcomes

By the end of this course, students will be able to perform statistical analyses that involve critical assessment of data and model outcomes to ensure that they are applying the most appropriate statistical methods. Specifically, students will be able to define the characteristics and properties of the multivariate normal distribution. Characterize the concept of interdependency among multidimensional data through correlation and the meaning of independence. Perform Principle Components Analysis, Factor Analysis, Cluster Analysis and Discriminant Analysis for multidimensional data sets, regression and interpret the results of these analyses.

Statistical Software R

The methods discussed in this course shall be illustrated using the statistical software R, for which one (or two) classes every two weeks shall be devoted to learning such implementations. All students are required to install on their personal laptops, the latest version of the software “R” from the website, https://www.r-project.org/, along with the user interface “Rstudio” from the website https://www.rstudio.com/. Students shall be informed when to bring their laptops to class to discuss these implementations. Some homework problems shall also be based on these discussions.

Recommended Course Material

**Grading**

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Assignments (Bi-weekly)</td>
<td>60%</td>
</tr>
<tr>
<td>Midterm Exam</td>
<td>25%</td>
</tr>
<tr>
<td>Final Exam</td>
<td>15%</td>
</tr>
</tbody>
</table>

**Attendance:** Lecture attendance is highly recommended. It is the student’s responsibility to make up on missed material due to class absence.

The following letter grades based on absolute percentage score will be guaranteed:

- 93% and above: A
- 90%-92.99%: A-
- 87%-89.99%: B+
- 83%-86.99%: B
- 80%-82.99%: B-
- 77%-79.99%: C+
- 73%-76.99%: C
- 70%-72.99%: C-
- 67%-69.99%: D+
- 60%-66.99%: D

**Homework Assignments:**

Problem sets will be assigned roughly bi-weekly during the entire semester. These are designed to give students practice in applying the quantitative techniques learned in class. While conceptual discussions among students are encouraged, however homework assignments are required to be individual work.

**Exams**

This course will have a midterm and a final exam. The exams will be open book, open notes and will be mainly designed to see if: 1) the students have understood the primary concepts covered in the course (see the student learning outcomes); 2) have read the material; and 3) can apply what they have learned in (somewhat) new situations.

**Use of Electronic Devices in Class**

No electronic devices other than computers are to be used during class. Cell phones should be turned off. Computers are only to be used when you are instructed to do so and only for example purposes. You may be asked to leave the classroom if you violate this policy.

**For Persons with Disabilities**
Reasonable accommodations are available for students with a documented disability. If you have a disability and may need accommodations to fully participate in this class, please either visit or call the Access Center (Washington Building 217; 509-335-3417) to schedule an appointment with an Access Advisor. All accommodations MUST be approved through the Access Center.

Campus Safety

Students should familiarize themselves with the following links regarding safety at WSU:


Academic Integrity

Students are encouraged to visit the Washington State University Academic Integrity Program website at: http://academicintegrity.wsu.edu/.

This university and this instructor take academic integrity seriously. Violations including, but not limited to, cheating and plagiarism, may result in penalties ranging from losing points on an assignment to failure in the course. Do not copy answers for assignments and exams. Such behavior creates a blatant misrepresentation of one’s achievements, and it damages the very foundation of academia.

Course Schedule and Outline

The schedule of topics is not fixed. Enough time will be spent on each to ensure students have a solid grounding in one topic until we move to another.

- **Topic 1 – Week 1 and 2:**

- **Topic 2 – Week 3 and 4:**
  The Multivariate normal distribution. Some properties of the MVN. More on Correlation Coefficients. Inference on Means.

- **Topic 3 – Week 5:**
  Inference on Means.

- **Topic 4 – Weeks 6 and 7:**
  Multivariate analysis of variance (MANOVA)

- **Topic 5 – Weeks 8 and 9:**
  Multiple and multivariate regression. Principal Component Analysis (PCA) as a dimensional reduction method.

- **Topic 6 – Weeks 11 and 12:**
  PCA continued. Factor analysis.

- **Topic 7 - Week 13**
  Classification and Discriminant Analysis.
• **Topic 8 – Weeks 14 and 15**
  Clustering. Multidimensional Scaling.

• **Topic 9 – Week 16:**
  Advanced Topics