King Fahd University of Petroleum and Minerals
Department of Mathematics and Statistics
STAT-503: Probability and Statistics for Data Science (Term 222)

Instructor: Mohammad Saleh
Office: 5-312
Phone: 013-860-4410
E-mail: mohfarah@kfupm.edu.sa
Office Hours: will be announced later

Course Description: Selected topics from Probability theory, Statistical Inference, and Information Theory for Data Science with an emphasis on the implementation using statistical software, toolboxes, and libraries like R, NumPy, SciPy, Pandas, and Statsmodels. Topics include Probability; Conditional Probability; Bayes' Theorem; Random variables; Discrete and Continuous Distributions; Central Limit Theorem; Point Estimation MLE and MAP; Confidence Interval Estimation; Hypothesis Testing; Non-parametric Statistics; Synthetic Data; Entropy, Mutual Information; Information Gain.

Textbook: Matloff, N. Probability and Statistics for Data Science Math + R + Data, CRC Press 2019

Assessment*

Activity	Weight
Classwork (home works, assignments, attendance, bonuses, etc.)	10%
Quizzes	10%
First Exam	15%
Second Exam	15%
Project	15%
Final Exam (Comprehensive)	35%

Week	Topics	Remarks
Week 1	Probability	General concept of probability Sample space and events
Week 2	Probability	Definition of probability Counting techniques Addition theorem
Week 3	Probability	Conditional probability Multiplication rule Independence of events Bayes' theorem
Week 4	Random variables	Definition of random variable with examples. Types: Discrete and Continuous. Probability distribution
Week 5	Probability distribution	Discrete Probability distributions, including: Binomial distribution, Hypergeometric distribution, Geometric distribution and Poisson distribution
Week 6	Probability distribution	Uniform distribution Exponential distribution
Week 7		Normal distribution Other continuous distributions.
Week 8	Estimation	Central Limit Theorem; Point Estimation MLE and MAP

Week 9	Interval estimation	Confidence Interval Estimation
Week 10	Hypothesis Testing;	Hypothesis Testing
Week 11	Hypothesis Testing;	Hypothesis Testing
Week 12	Non-parametric Statistics	Non-parametric Statistics
Week 13	Non-parametric Statistics	Non-parametric Statistics
Week 14	Synthetic Data; Entropy,	
Week 15	Mutual Information; Information Gain.	

Important Notes: .

Excuse: Only an excuse issued by ***Deanship of Student Affairs*** will be accepted for not attending a class, a quiz or an exam.

Blackboard: All contacts or announcements between the instructor and the students are supposed to be through Blackboard, so the student must check his Blackboard at least once a day.

Academic Integrity: All KFUPM policies regarding ethics and academic honesty apply to this course.

Cheating and Plagiarism: This course is composed of individual assignments. It is important that your individual assignment be completed with your own efforts instead of copying it from your fellow student. KFUPM instructors follow "***zero tolerance***" approach with regard to cheating and plagiarism. During examinations (quizzes, major exams, lab tests) cheating or any attempt of cheating by use of illegal activities, techniques and forms of fraud will result in a ***grade of F*** in the course along with reporting the incident to the higher university administration.

Important Notes:

- In accordance with University rules, 20% unexcused absences will automatically result in a grade of DN.
- Attendance on time is very important. Mostly, attendance will be checked within the first five minutes of the class. Entering the class after that, is considered as one late, and every two times late equals to one absence.
- Use of mobile is banned during the class