Instructor
Saurav Kumar (915) 747 6329 skumar2@utep.edu.

Course Objectives
At the completion of this class, students should be able to:
1. Conduct exploratory data analysis including measures of central tendency and dispersion, histograms, box plots, cumulative density function plots, Q-Q plots, and stem and leaf plots.
2. Understand assumptions and limitations for hypothesis tests
3. Design, test assumptions, conduct, and interpret parametric and nonparametric statistical tests
4. Develop regression models
5. Test for trends using regression and nonparametric methods
6. Handle censored data.

The course is designed to introduce students to tools and references available. Students are expected to develop deep understanding in at least one topic that may be most relevant to their research or work.

Modules
1. Introduction to data analysis and wrangling using python and pandas
2. Basics statistics
   a. Distributions
   b. Parametric and non-parametric
   c. Describing uncertainty
   d. Hypothesis testing
   e. Testing two groups
      i. Matched pair
   f. Testing several independent groups
3. Regression
   a. Correlation
   b. Simple linear regression
   c. Alternative methods for regression
   d. Multiple linear regression
   e. Discrete regression
4. Trend analysis
5. Introduction to bootstrapping techniques
6. Handling censored data

A more detailed schedule will be available closer to the class start date.
**Grading**

**Homework (40% of the grade)**

Students will be required to work with a dataset after each module and write a memo describing the analysis.

Memos will be graded on writing style (10%), analysis of data (60%), and conclusions derived (30%).

**Analysis Paper (30% of the grade)**

Students, in groups of 2 or individually, will prepare a paper based on any dataset of interest they have analyzed. The dataset may be taken form a previously published study. The analysis paper will be due on the last day of the class.

Papers for undergraduate students will be graded on the writing style (10%), analysis of data (60%), and conclusions derived (30%).

Papers developed by graduate students should be submission ready to a journal of their choice. That will require writing a detailed introduction to the topic and the problem, describing the dataset and methodology, describing the results, and finally a discussion and conclusion. Papers for Graduate students will be graded taking all the section of the paper into account.

**Midterm Exam (10% of the grade)**

Midterm will be a 1-hour exam. All exam questions will be multiple choice with little computation needed. Students may bring a cheat-sheet (handwritten, one side, letter size paper) to the exam.

**Final Exam (20% of the grade)**

Final will be a comprehensive 2-hour exam. All exam questions will be multiple choice with little computation needed. Students may bring a cheat-sheet (handwritten, one side, letter size paper) to the exam.

**Final grade**

<table>
<thead>
<tr>
<th>Final Average, %</th>
<th>Course Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥90</td>
<td>A</td>
</tr>
<tr>
<td>80-89</td>
<td>B</td>
</tr>
<tr>
<td>70-79</td>
<td>C</td>
</tr>
<tr>
<td>60-69</td>
<td>D</td>
</tr>
<tr>
<td>&lt; 60</td>
<td>F</td>
</tr>
</tbody>
</table>

**References**


5. Statistics for Censored Environmental Data Using Minitab and R 2nd Edition by Dennis R. Helsel. (reference text; not required)
7. A Primer on Scientific Programming with Python (Texts in Computational Science and Engineering) 4th ed. by Hans Petter Langtangen. (reference text; not required)

Software
Students may choose to use any software for analysis including MS Excel and Minitab. Instruction will be based on the following software

- R
- Python
- Anaconda (Python and R installer with other libraries)
- JMP (Spreadsheet based statistical software)

Everything, except JMP, is available free of cost. Students who do not have programming background may find JMP easier to use.

It will be beneficial for students to go through https://www.codecademy.com/learn/python and http://tryr.codeschool.com/ or any other online tutorial that introduces basics of writing code in Python and R before the first day of the class.

Other notes
Plagiarism or cheating will not be tolerated; anyone caught cheating will receive a failing grade for the class.

If you have a disability and need classroom accommodations, please contact The Center for Accommodations and Support Services (CASS) at 747-5148, or by email to cass@utep.edu, or visit their office located in UTEP Union East, Room 106. For additional information, please visit the CASS website at www.sa.utep.edu/cass.