DS-405 Experimental Design for Data Science

Credit Hours 3-0

Pre-requisites: NONE

Objectives: Data Science project exists the planning, design and execution of experiments. Experimental design is the process of carrying out research in an objective and controlled fashion so that precision is maximized and specific conclusions can be drawn regarding a hypothesis statement. Generally, the purpose is to establish the effect that a factor or independent variable has on a dependent variable

Course Contents

Principles of experimental design, Layout analysis and related efficiency of completely randomized, randomized complete block, Latin square, Cross-over and Greco Latin square designs, Estimation of missing observations in three basic designs, Fixed, random and mixed effect models in the basic designs, Factorial experiments , Multiple comparisons, Effect of violation of assumptions underlying ANOVA and transformation of data, Analysis of covariance up to two covariates.Practically structuring lay-out plan of basic experiments techniques, collection of data, estimation of parameters and statistical analysis on collected data. Response surface methodology and its optimization

Outcomes: At the end of the course, attendees should be able to:

- Describe some of the factors affecting reproducibility and external validity.
- Conduct the different types of formal experimental design
- Optimize the experimental design.

Textbook Montgomery, Douglas C. Design and analysis of experiments. John wiley & sons, 2017.

Reference Books:

Quinn, Gerry P., and Michael J. Keough. Experimental design and data analysis for biologists. Cambridge university press, 2002.

Gonzalez, Richard. Data analysis for experimental design. Guilford Press, 2009.

Diamond, William J. Practical experiment designs: for engineers and scientists. John Wiley & Sons, 2001.

Weekly Breakdown		
Week	***Section***	***Topics***
1	1.1-1.4	Introduction and basic strategies
2	3.2-3.3, 3.3.1	The Analysis of Variance, Analysis of the Fixed Effects Model Decomposition of the Total Sum of Squares
3	3.3.3, 3.4	Estimation of the Model Parameters, Unbalanced Data, Model Adequacy Checking
4	4.1-4.2	The Randomized Complete Block Design, The Latin Square Design
5	5.1-5.2	Factorial Designs,Basic Definitions and Principles, The Advantage of Factorials
6	5.3	The Two-Factor Factorial Design
7	5.4	The General Factorial Design
8	5.5	Fitting Response Curves and Surfaces
9	**Mid Semester Exam**	
10	5.6	Blocking in a Factorial Design
11	6.1-6.2	2^2 Design
12	6.3	2^3 Design
13	6.4-6.5	General 2k Design, A Single Replicate of the 2k Design
14	6.7	2^k Designs are Optimal Designs
15	11.1-11.2	Introduction to Response Surface Methodology. The Method of Steepest Ascent
16	11.3-11.4	Analysis of a Second-Order Response Surface, Location of the Stationary Point
17		Revision
18	**End Semester Exam**	