

# Ekaagra: Programs Offered

**April 2021**

**Statistical Foundations  
of  
Data Science (SFDS)  
Series**

# SFDS-01: Basic Statistics with R

9-hr online course via Google Meet

All exercises will be conducted using R, thus SFDS-1 is very fundamental

## ➤ Part-1: Introduction to R

- ❖ History of R
- ❖ Installing R
- ❖ Basic Operations in R – Objects, Functions, Plots
- ❖ Basic Statistics in R

## ➤ Part-2: Probability and Probability Distributions, 2 hr

- ❖ Probability: Addition and multiplication rules, Bayes' Theorem
- ❖ Probability Distributions: Binomial, Poisson and Normal

## ➤ Part-3: Sampling Distributions and Hypothesis Testing

- ❖ T-Distribution
- ❖ Confidence Intervals
- ❖ Type-I and Type-II errors
- ❖ T-tests and ANOVA

## ➤ Part-4: Hypothesis Testing (cont'd)

- ❖ F-Test and Homogeneity of Variance
- ❖ Chi-Squared
- ❖ Non Parametric Tests – can be elaborated separately if required

# SFDS-02: Data Visualization and Time Series

9-hr online course via Google Meet

The entire program will be done in the R Programming Language

## ➤ Part-1: Three Distributions and Basic Data Analysis

- ❖ Normal Distribution and Descriptive Statistics
- ❖ Binomial and Poisson Distributions, and their descriptive statistics

## ➤ Part-2: The Visual Presentation of Data

- ❖ Histogram, Stem and Leaf Plot, Probability Plotting, Bar Chart, Pareto, Pie Chart, Dot Plot,
- ❖ Box Plot, Scatter Plot, Matrix Plots, Chernoff-Flury Faces

## ➤ Part-3: Elements of Time Series-A

- ❖ Basics of Forecasting
- ❖ Time Series Graphics

## ➤ Part-4: Elements of Time Series-B

- ❖ Time Series Decomposition
- ❖ Time Series Modeling: Exponential Time Smoothing (ETS) class incl. Holt-Winters
- ❖ Time Series Modeling: ARIMA class of models
- ❖ Prediction using above two classes

# SFDS-03: Statistical Learning – Regression

8-hr online course via Google Meet

The entire program will be done in the R Programming Language

- **Part-1: Statistical Learning and Linear Regression, 2 hr**
  - ❖ Introduction to Statistical Learning – Supervised and Unsupervised
  - ❖ Regression vs Classification
  - ❖ Method of Least Squares
  - ❖ Linear Regression
- **Part-2: Linear and Multiple Regression, 2 hr**
  - ❖ Model Adequacy for Linear Regression
  - ❖ Multiple Regression
  - ❖ Handling Qualitative Predictors
- **Part-3: Classification Techniques, 2 hr**
  - ❖ Logistic Regression
  - ❖ Linear Discriminant Analysis
  - ❖ K-Nearest Neighbors
- **Part-4: Resampling Methods**
  - ❖ Cross Validation
  - ❖ Bootstrap

# SFDS-04: Statistical Learning – Advanced Topics

8-hr online course via Google Meet

The entire program will be done in the R Programming Language

## ➤ Part-1: Model Selection and Regularization

- ❖ Subset Selection
- ❖ Shrinkage Methods: Ridge Regression and Lasso
- ❖ Dimension Reduction Methods: Principal Components and Partial Least Squares

## ➤ Part-2: Tree Based Methods, 2 hr

- ❖ Classification and Regression Trees
- ❖ Bagging and Random Forests

## ➤ Part-3: Support Vector Machines, 2 hr

- ❖ Maximum Marginal Classifier
- ❖ Support Vector Classifier

## ➤ Part-4: Unsupervised Learning

- ❖ Principal Components Analysis (PCA)
- ❖ K-Means Clustering

Do you feel utterly lost and miserable when you see data?  
Do you think that statistics is too important to be left to statisticians?  
Have you ever got the feeling that you've been had by some shyster slicing  
and dicing data in a way that you think isn't quite kosher?

If yes, do tune in to

# STATISTICS FOR *Non-Statisticians*

*Sat 7<sup>th</sup>* NOVEMBER 2020

*1800-2000 IST, followed by a short Q&A session*

This will be a live session on Google Meet, and will cost you  
a quintal of gold.

Well, not really, we're happy with a thousand Indian rupees payable  
in advance via G-Pay. You can pay via the QR code in this picture or  
transfer to this UPI ID via G-Pay: [menonramdas8652@icici](mailto:menonramdas8652@icici)

Do send me a mail after you fork out your cash!

**RAMDAS MENON**

[www.ekaagra.com](http://www.ekaagra.com)  
[ramdas@ekaagra.com](mailto:ramdas@ekaagra.com)



# Agenda: Statistics for Non Statisticians

- **General Introduction**
- **Slicing and Dicing**
- **Descriptive Statistics**
- **Probability**
- **Distributions**
- **Confidence Intervals**
- **Statistical Inference: the concept of Hypothesis Testing**
- **Putting it together**



# Design of Experiments (DOE) Series

# Design of Experiments (DOE) – Basic

- **The Three Pillars of DOE: Replication, Randomization and Blocking**
- **The One Factor At a Time (OFAT) Approach and its limitations**
- **Factorial Experimentation: basic concepts**
- **Fractional Factorials: getting more with less**
- **Resolution of designs**
- **Audience: R&D, Engg, Quality, Production**
- **Duration : 3 days**
- **Prerequisite: Basic and Intermediate Statistical Method Modules**

# Design of Experiments (DOE) – Advanced

- **Response Surface Methods**
  - ❖ Box-Wilson
  - ❖ Box-Behnken
- **3 level factorials – overview**
- **Screening Designs:**
  - ❖ Plackett-Burman
  - ❖ Taguchi
- **Taguchi's Robust Designs**
- **RCBD-s and BIBD-s**
- **Audience: R&D, Engg, Quality, Production**
- **Duration : 2 days**
- **Prerequisite: Basic DOE Module**

# Design of Experiments (DOE) – Mixtures

- **Mixtures vs Factorial Designs**
- **Synergism and Antagonism**
- **Transfer Functions**
- **Two Component and Three Component Mixtures**
- **Triangular Diagrams**
- **Mixture Types:**
  - ❖ Simplex Lattice
  - ❖ Simplex Centroid
  - ❖ Extreme Vertices
- **Audience: R&D, Engg, Quality, Production**
- **Duration : 2-4 days, depending on whether Mixtures are included**
- **Prerequisite: Basic DOE Module**

# Statistical Quality Control (SQC) Series

# Statistical Quality Control (SQC) – Basic

- **Shewhart and his philosophy of control charts**
- **Charts for Continuous Data**
  - ❖ X-Bar R and X-Bar S charts
  - ❖ I-MR charts
- **Charts for Discrete Data**
  - ❖ P and NP charts for defectives – Binomial
  - ❖ C and U charts for defects – Poisson
- **Process Capability from Control Charts**
- **Gage Capability Studies (Optional)**
- **Audience: Quality, Production**
- **Duration : 3 days**
- **Prerequisite: Basic and Intermediate Statistics modules**

# Statistical Quality Control (SQC) – Advanced

- **Quick Revision of Charts for Continuous Data**
- **Quick Revision of Charts for Discrete Data**
- **Other types**
  - ❖ Cumulative Sum (Cusum) Charts
  - ❖ Exponentially Weighted Moving Average (EWMA) charts
- **Process Capability from control charts**
- **Integrating Engineering Process Control (EPC) with Statistical Process Control**
- **Audience: Quality, Production**
- **Duration : 2 days**
- **Prerequisite: SQC Basic Module**

**Mathematical Foundations  
of  
Data Science (MFDS)  
Series**



# MFDS: Some Thoughts

- **We will roll this out by September 2021**
- **This will cover the following subjects**
  - ❖ Linear Algebra
  - ❖ Multivariate Calculus
  - ❖ Probability
- **Each subject will be covered in 4 lectures of 2 to 2.5 hr each**



# Other Programs

# Statistical Thinking

- **Concept of process**
- **Concept of variation**
- **Organization viewed as a system of interacting processes**
- **Importance of statistics in day-to-day work**
- **The two types of errors in decision making – Type I and Type II**
- **Statistical Process Control**
- **A brief look at statistical tools and methods**
- **Audience: all functions in a company**
- **Duration : One day**
- **Prerequisite: None**

# Statistical Methods – Basic

- **Types of Data and Graphical representation**
- **Characterization of Data:**
  - ❖ Measures of Central Tendency and Dispersion
  - ❖ Skewness and Kurtosis
- **Probability : basic ideas – sample space, counting rules, joint probability**
- **Distributions: Binomial and Gaussian (Normal) – practical applications**
- **Sampling Distributions**
- **Confidence Intervals**
- **Audience: all functions in a company**
- **Duration : 3 days**
- **Prerequisite: None**

# Statistical Methods – Intermediate

- **Probability: advanced concepts such as**
  - ❖ conditional probability,
  - ❖ Bayesian analysis
- **Distributions other than Normal**
  - ❖ Discrete – Hypergeometric, Poisson
  - ❖ Continuous – Exponential, Weibull
- **Probability plotting**
- **Hypothesis Testing for Location, Spread and Independence**
  - ❖ 1-sample, 2-sample and paired t tests
  - ❖ Analysis of Variance (ANOVA), F-test and Homogeneity of Variance
  - ❖ Chi-squared test
- **Simple Linear Regression**
- **Gage Capability Process Capability**
- **Audience: Planning, Quality, Engg , Operations and R&D functions**
- **Duration : 3-4 days**
- **Prerequisite: Basic Statistics module**

# Statistical Methods – Advanced

- **Power and Sample Size Calculations**
- **Multiple Regression**
- **Non Parametric (Distribution free) tests**
- **Audience: Engg, Quality and R&D**
- **Duration : 2 days**
- **Prerequisite: Basic and Intermediate Statistics modules**

# Multi Variate Analysis (MVA)

- **Multi Variate Normal Distribution**
- **Partial and Multiple Correlation Coefficients**
- **Hotelling T2 statistic and T2 Control Chart**
- **Principal Component Analysis (PCA)**
- **Factor Analysis**
- **Multiple Analysis of Variance (MANOVA)**
- **Multivariate Analysis of Covariance (MANCOVA)**
- **Logistic Regression**
- **Correspondence Analysis**
- **Audience: R&D, Engg, Quality, Sales & Marketing**
- **Duration : 2 days**
- **Prerequisite: Basic and Intermediate Statistical Methods Modules**

# Reliability Engineering

- **Basic Terminology such as MTBF, MTTF, MTTR, Hazard rate and Availability**
- **Distributions: Normal, Exponential, and Weibull**
- **Statistical Inference**
- **Probability plotting**
- **Analysis of Life Data – complete, singly censored and multiply censored**
- **Reliability Modeling**
- **Calculation of reliability parameters**
- **Highly Accelerated Life Testing (HALT)**
- **Highly Accelerated Stress Screening (HASS)**
- **Audience: R&D, Engg, Quality**
- **Duration : 2 days**
- **Prerequisite: Basic and Intermediate Statistical Methods Modules**



# Process Mapping

- **The rationale for Process Mapping**
- **SIPOC**
- **Flow Charting and Process Variables Mapping**
- **Swim Lanes or Cross Functional Maps**
- **Value Stream Mapping**
- **Audience: Sales & Marketing, Planning, Engg, Quality, R&D**
- **Duration: 2 days**
- **Prerequisite: None**

# Seven QC Tools – Traditional

## ➤ **TRADITIONAL**

- ❖ Check Sheet
- ❖ Pareto
- ❖ Cause and Effect (Ishikawa) Diagram
- ❖ Stratification
- ❖ Histogram
- ❖ Scatter Diagram
- ❖ Control Chart

➤ **Audience: all functions in a company**

➤ **Duration: 2 days**

➤ **Prerequisite: None**

# Seven QC Tools - Modern

## ➤ MODERN

- ❖ Affinity Diagram
- ❖ Interrelationship Digraph
- ❖ Tree Diagram
- ❖ Prioritization Matrix
- ❖ Matrix Diagram
- ❖ Process Design Program Chart (PDPC)
- ❖ Activity Network Diagram

➤ **Duration : 2 days**

➤ **Audience: all functions in a company**

➤ **Prerequisite: None**

# Monte Carlo Simulation

- **Quantitative estimation of risks involved in any process**
- **The range of applicability could range from manufacturing and marketing to financial planning and project management**
- **Used for predictive modeling, forecasting, simulation and optimization**
- **We recommend R Programming Language.**
- **Hands on solutions of your day to day problems can be taken up separately**
- **Audience: Sales & Marketing, Planning, Engg, Quality, R&D**
- **Duration: 2 days**
- **Prerequisite: None**