# ADVANCED STATISTICS ( Week 1)

greatlearning
Power Ahead

# DSBA CURRICULUM DESIGN

## FOUNDATIONS

Python for Data Science

Statistical Methods for Decision Making

## CORE COURSES

Advanced Statistics(Week-1/5)

Data Mining

Predictive Modelling

Machine Learning

Time Series Forecasting

Data Visualization

SQL

## DOMAIN APPLICATIONS

Financial Risk Analytics

Marketing Retail Analytics

greatlearning
Power Ahead

# LEARNING OBJECTIVE OF THIS MODULE

- ANOVA
- EDA
- PCA

# LEARNING OBJECTIVES OF THIS SESSION - APPLICATION OF HYPOTHESIS TESTING

- Application of One way ANOVA
- Application of two way ANOVA

greatlearning
Power Ahead

# TRY ANSWERING THE FOLLOWING

- What do we compare in ANOVA? Is it Mean or Variance?
- How many minimum different sets of sample data is required for implementing ANOVA?
- What is the Null Hypothesis(Ho) of one way ANOVA given there are 3 different populations?

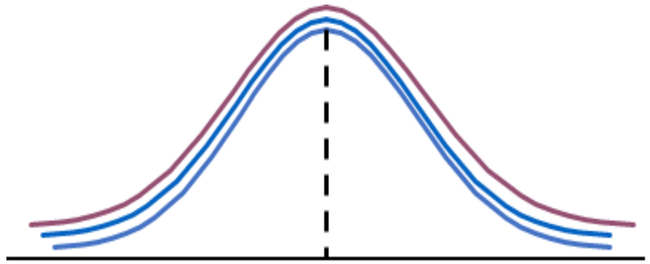
# BROAD OVERVIEW

## One Way ANOVA

$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$

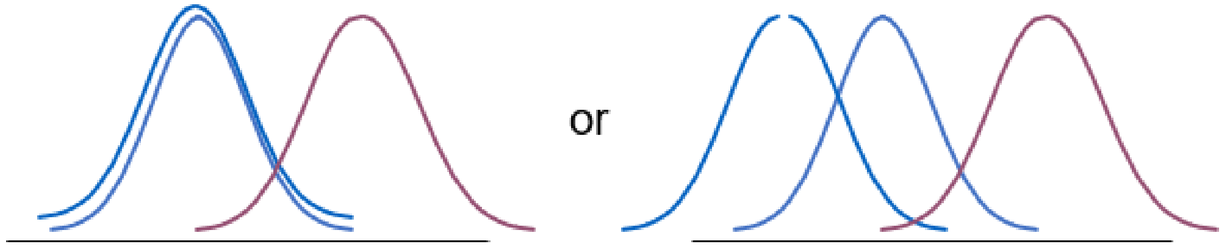
$H_1 : \text{Not all } \mu_j \text{ are equal}$

The Null Hypothesis is True
All Means are the same



$\mu_1 = \mu_2 = \mu_3$

The Null Hypothesis is NOT true
At least one of the means is different



$\mu_1 = \mu_2 \neq \mu_3$ or $\mu_1 \neq \mu_2 \neq \mu_3$

# Analysing the Effectiveness of COPD Drugs Through ANOVA

UniChronic Obstructive Pulmonary Disease, is a chronic lung disease that makes it hard to breathe because less air flows in and out of the airways in lungs. When you're getting less air, less oxygen gets into body tissues and it gets harder to get rid of the waste gas carbon dioxide. This results in shortness of breath during everyday activities. People with COPD can
experience fatigue, chronic cough and frequent respiratory infections as well. COPD doesn't just have a physical impact—living with chronic disease also can affect mental health.

**Use of ANOVA**-To Analyse the effectiveness of 3 COPD drugs Advair, Symbicort and Spiriva, and their combinations, tests were conducted in terms of patients' length of Stay in the hospital.Samples are collected for each drug, and length of stay in the hospital for different patients are the observations contained in these samples. Finally the mean lengths of stay across drugs are compared  to comment on the effectiveness of the drugs. Lower the mean stay, better the drug!

https://support.sas.com/resources/papers/proceedings17/1031-2017.pdf

# ANOVA's GDP  Connection- Importance of Agriculture ?

Talking about India, agriculture is the most important sector of our Economy. Indian agriculture sector accounts for 18 per cent of India's gross domestic product (GDP) and provides employment to 50% of the countries workforce. Similarly, In America agriculture contributes to $1.053 Trillion to US  Gross GDP.

**Use of ANOVA-** The goal of research on fertilizer rate is to determine the amount of fertilizer needed to achieve a commercial crop yield with sufficient quality that is economically acceptable for the grower. The researcher applied a range of fertilizer rates thought to capture the likely extent of possible crop yield responses. Finally, a comparison between the yield samples belonging to different fertilizer rates helps decide which fertilizer rate is better.

Reference: https://edis.ifas.ufl.edu/pdffiles/SS/SS54800.pdf

greatlearning
Power Ahead

Let’s Learn Together – A Unique Platform for Peer to Peer Learning

## Next Week’s Theme:

## Mind-Map of Advance Statistics Course

**Reference Link**-https://www.mindmapping.com/

## Benefits of Creating Mind-Map:

- Quick Revision of the course
- This document will prove very handy later in the course
- Opportunity to show your creativity

## What all can be discussed in a Discussion forum?

- Analytical Concepts
- Issues in Code
- Real Time/Industry Examples

greatlearning
*Power Ahead*

# Data Science @ Work

Apply **Data Science at your workplace** to gain some instant benefits:

- Get noticed by your management with your outstanding analysis backed by data science.
- Create an impact in your organization by taking up small projects/initiatives to solve critical issues using data science.
- Network with members from the data science vertical of your organization and seek opportunities to contribute in small projects.
- Share your success stories with us and the world to position yourself as a subject matter expert in data science.

# Case Study-1 (GolfBall)

Sporting goods manufacturing company wanted to compare the distance travelled by golf balls produced using four different designs. Ten balls were manufactured with each design and were brought to the local golf course for the club professional to test. The order in which the balls were hit with the same club from the first tee was randomized so that the pro did not know which type of ball was being hit. All 40 balls were hit in a short period of time, during which the environmental conditions were essentially the same. The results (distance travelled in yards) for the four Designs are stored in Golfball.csv

At the 0.05 level of significance, is there evidence of a difference in the mean . distances travelled by the golf balls with different designs?

greatlearning
Power Ahead

# Case Study-2 (Dialysis)

The dose for effective dialysis depends on duration of treatment and weight gain between treatments. To study the effects of these two factors on the number of days hospitalized attributable to kidney failure, a random sample of 10 patients per group undergone treatment in a large dialysis facility was obtained. Treatment duration was at 3 levels: short, mid and long. Average weight gain between treatments during the year was categorized into three levels: mild, moderate and severe.

greatlearning
Power Ahead

# ANY QUESTIONS

greatlearning
*Power Ahead*

# HAPPY LEARNING