

ADVANCE STATISTICS (Week 1)

This file is meant for personal use by suchirjoon05@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

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

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

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(H_0) of one way ANOVA given there are 3 different populations?



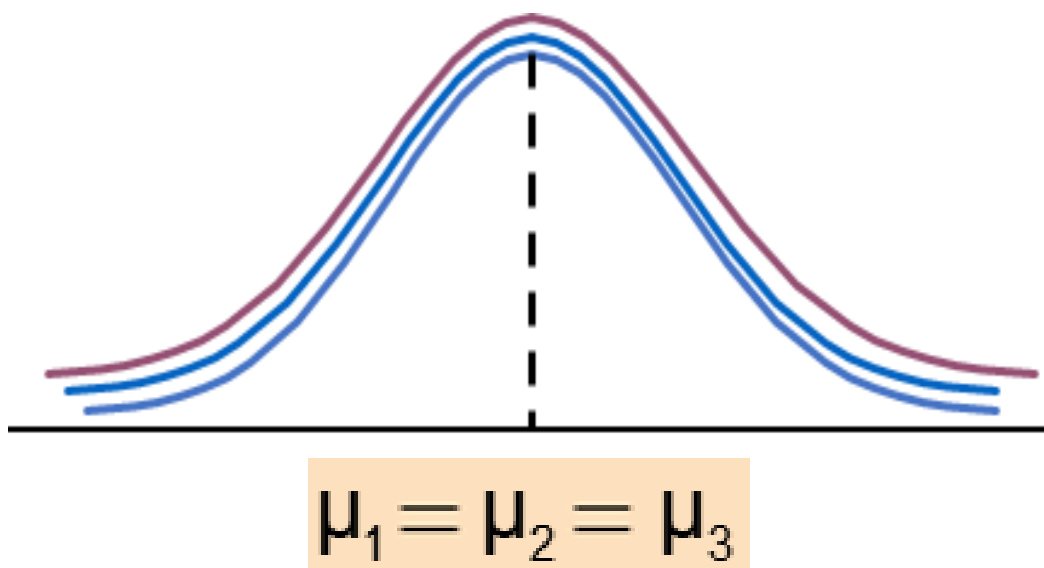
BROAD OVERVIEW

One Way ANOVA

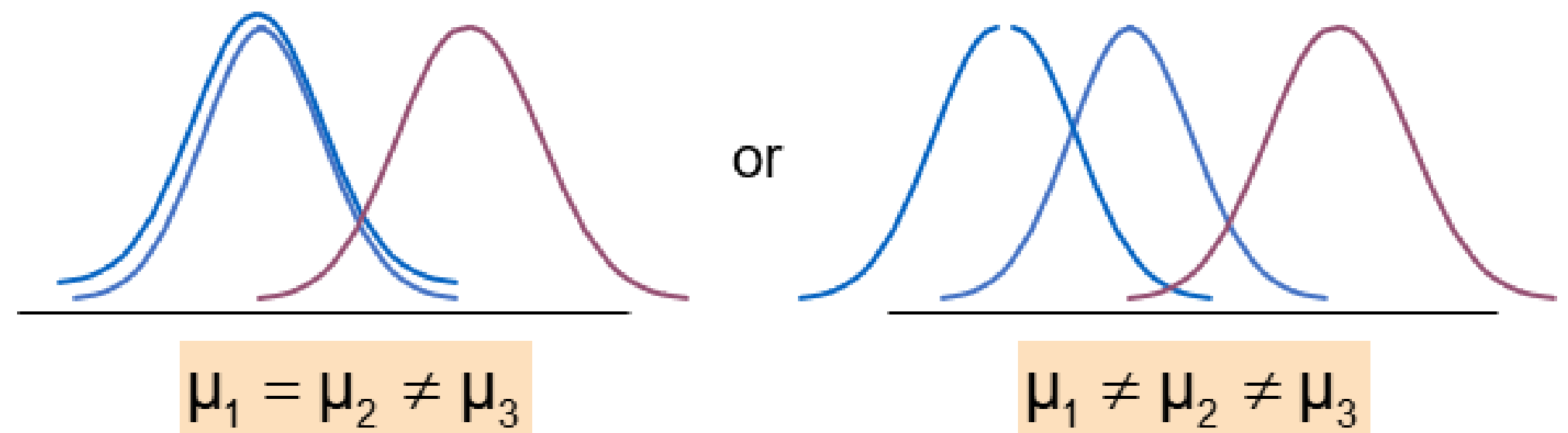
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_c$$

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

The Null Hypothesis is True
All Means are the same



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



Analysing the Effectiveness of COPD Drugs Through ANOVA

Chronic 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!



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/pdf/files/SS/SS54800.pdf>



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

Next Week's Theme:

EDA and It's Real Time/Industry Applications



Benefits of Peer to Peer Learning:

- ❖ Active Learning
- ❖ Gain a Deeper Understanding
- ❖ Feel More Comfortable
- ❖ Personalized Learning Experience

What all can be discussed in a Discussion forum?

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

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?

Case Study-2 (Diet)

The Health Company, which provides various diet plans to lose weight conducted a market test experiment conducted on both Gender (M, F) and tested three different level of diets (A, B, C). Each combination off Gender and Diet gave a different impact on the weight after following a diet for 6 weeks. Given is the dataset comprises of 78 observations which contains columns of Gender, Diet , Pre-weight and weight after 6 weeks. Perform ANOVA on the dataset and interpret the results.

Explain using Python.



ANY QUESTIONS



HAPPY LEARNING