

# STATISTICAL METHODS FOR DECISION MAKING (Week 1)



# GROUND RULES

- Come prepared for these sessions by watching the video lectures.
  - Concepts will be covered in the videos.
  - Hands-On Application will be covered in Mentor Sessions.
- Submit all assignments on time.
- Let's be punctual & respect each others time.



# DSBA CURRICULUM DESIGN

## FOUNDATIONS

Python for Data  
Science

---

Statistical Methods  
for Decision  
Making(Week-1/5)

## CORE COURSES

Advanced  
Statistics

---

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

- Descriptive Statistics
- Inferential Statistics
- Hypothesis Testing

# LEARNING OBJECTIVES OF THIS SESSION -

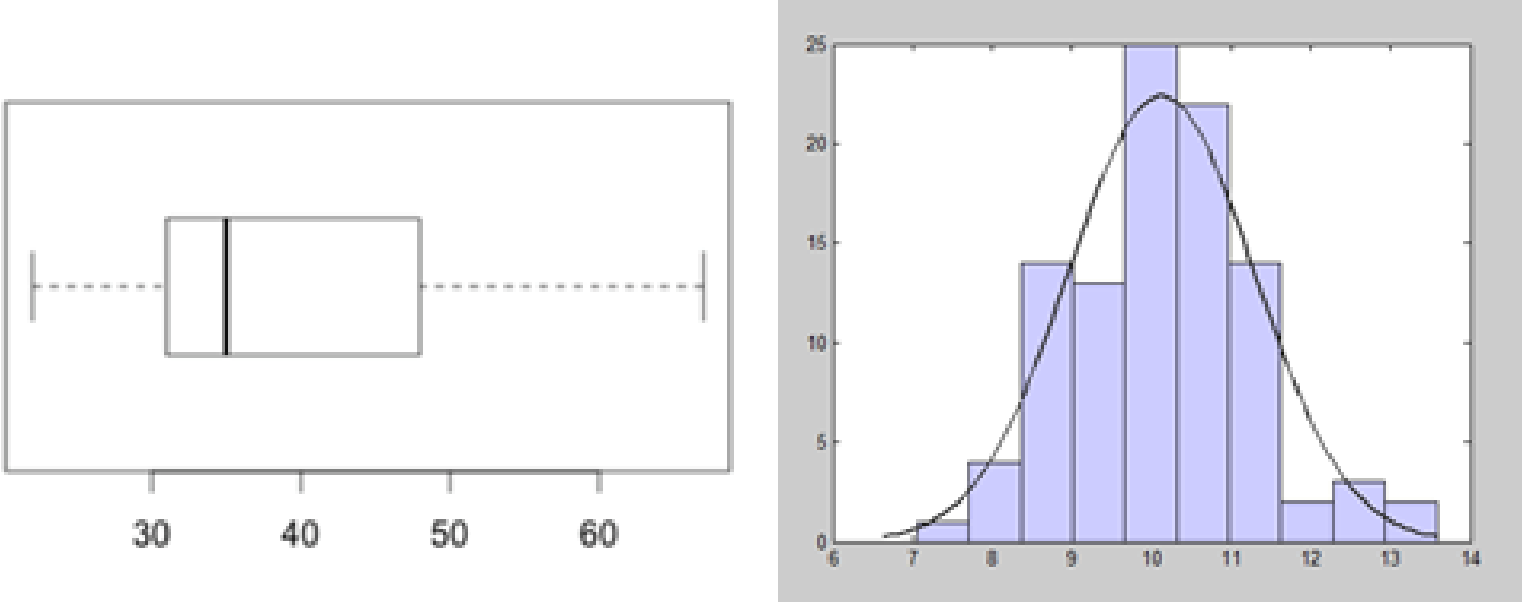
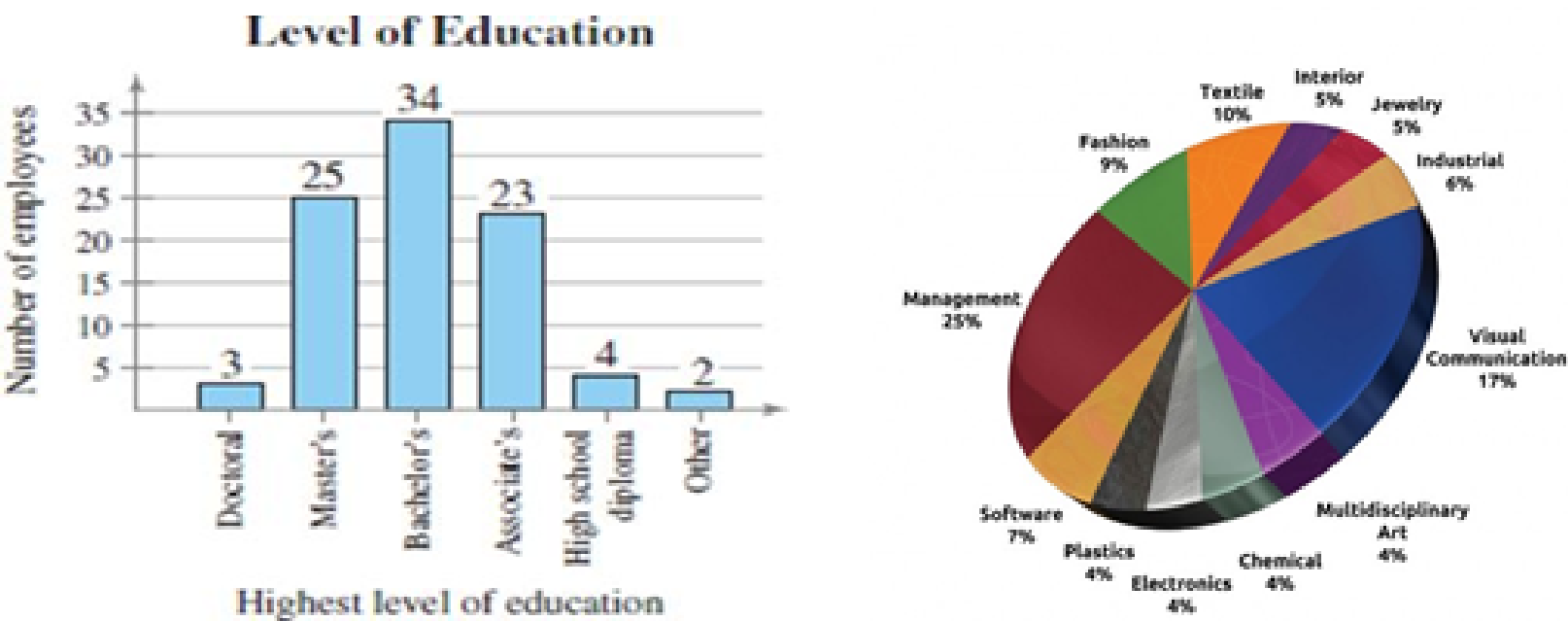
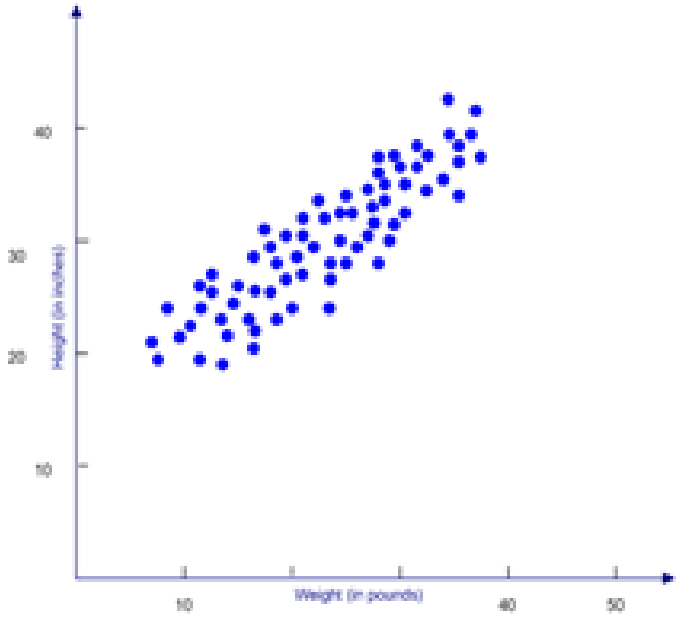
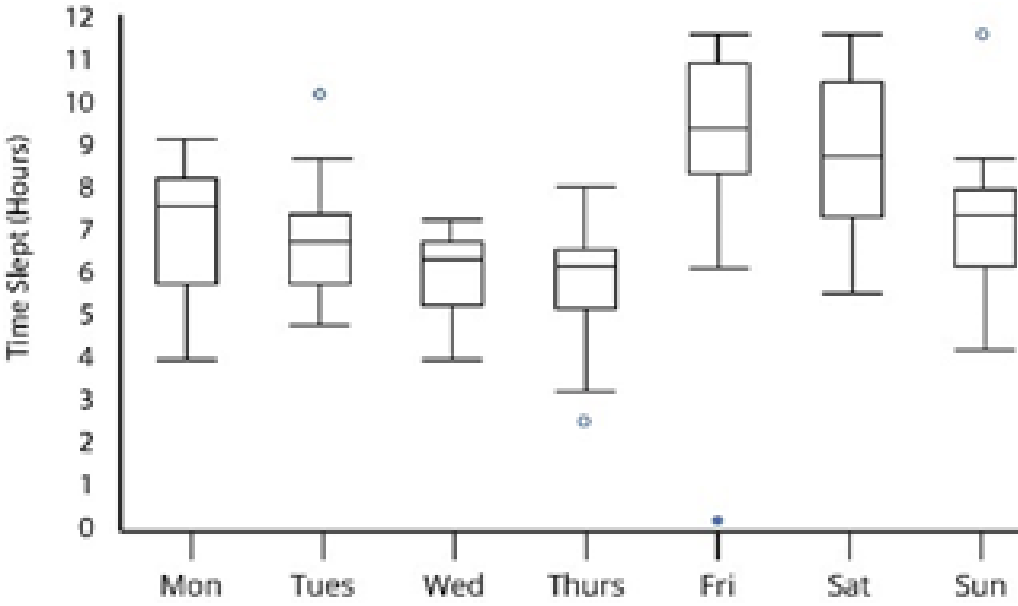
- Measures of Central Tendency
- Measures of Dispersion
- Graphical Display of Data
- Distribution and idea of Skewness
- Correlation

## TRY ANSWERING THE FOLLOWING

- Find Mean, Median, Mode and Range of the below data :-  
40,20,20,60,80,60,20,60
- What are the 5 points in a box-plot?
- What is the formula of Coefficient of Variation(CV)?



# BROAD OVERVIEW

	Numerical	Categorical
Univariate	<div><p>The univariate numerical section displays two charts. On the left is a box plot with a median around 35, an IQR from approximately 32 to 48, and whiskers extending from 25 to 65. On the right is a histogram with a normal distribution curve overlaid, centered at 10, with the x-axis ranging from 6 to 14 and the y-axis from 0 to 25.</p></div>	<div><p>The univariate categorical section contains two charts. The 'Level of Education' bar chart shows the number of employees for each education level: Doctoral (3), Master's (25), Bachelor's (34), Associate's (23), High school diploma (4), and Other (2). The pie chart shows the distribution of 15 different professions: Management (25%), Visual Communication (17%), Textile (10%), Fashion (9%), Software (7%), Plastics (4%), Electronics (4%), Chemical (4%), Multidisciplinary Art (4%), Industrial (6%), Jewelry (5%), Interior (5%), and others.</p></div>
Multivariate	<div><p>The multivariate numerical section features a scatter plot showing a positive correlation between 'Weight (in pounds)' on the x-axis and 'Height (in inches)' on the y-axis. The data points are clustered in the upper right quadrant, indicating that as weight increases, height also tends to increase.</p></div>	<div><p>The multivariate categorical section displays a box plot of 'Time Slept (Hours)' across the days of the week. The y-axis ranges from 0 to 12. The boxes show the median, quartiles, and range of sleep hours for each day, with outliers present for Tuesday, Thursday, and Sunday.</p></div>



# Application of Measures of Central Tendency

Which measure will be used in case of following scenarios ?

- When you try to search cool game app in play store, what measure will you look for? Is it the average rating?
- Suppose in your class there are 11 students and one of them is son of Bill Gate. How will you calculate average pocket money?
- How will you decide which color is most popular among citizens ?



# Application of Measures of Dispersion

For example, Rohit Sharma has a batting average of 47 runs per innings and the standard deviation is 30 runs. So more or less he scores between 17–77 runs on an average. Now if you look at scores of Shikhar Dhawan. He has an average of 45 runs per innings and his standard deviation is also 15 runs. So he scores somewhere between 30–60 runs on an average.

Now since both have nearly similar batting average, whom can we count on to score more consistently?

## Industry Application - Personality Assessment

In order for one to make meaningful statements about psychological events, the variable or variables involved must be organized, measured, and then expressed as quantities. Such measurements are often expressed as measures of central tendency and measures of variability.

In psychological testing, Descriptive statistics give a general picture of the scores in a given group. They include the measures of central tendency and the measures of variability. Central tendency involves different kinds of averages: the mean, median, and mode. Variability involves the standard deviation, which indicates how far scores in a group are likely to be from the average.

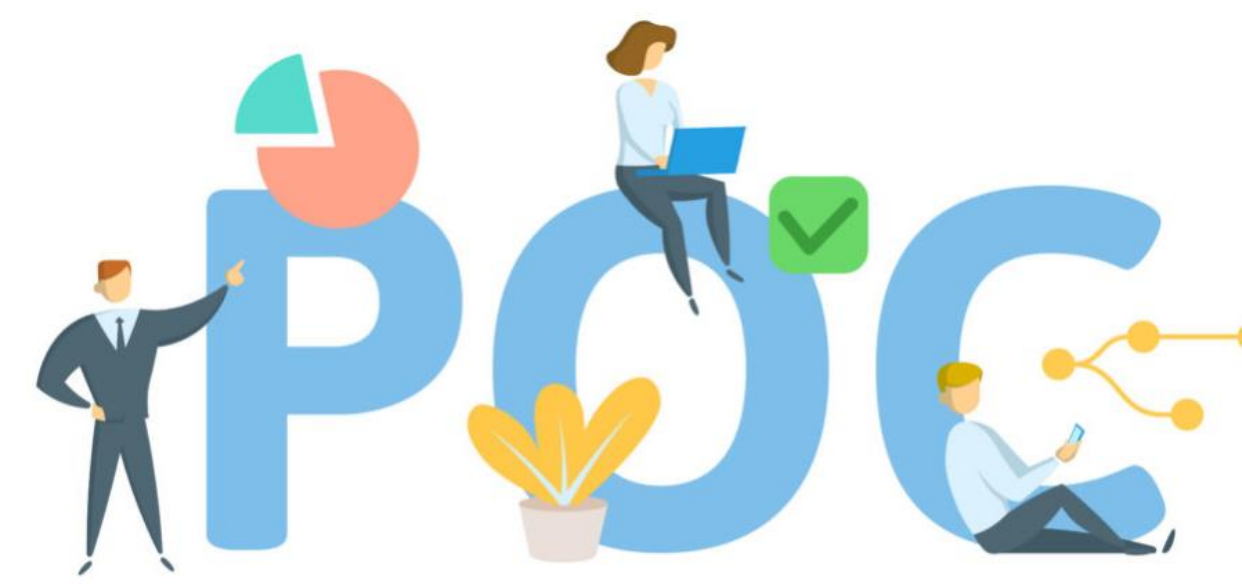


## CASE STUDY - HEALTH INSURANCE

Most companies are now recognizing the power of data in making crucial business decisions. For an Insurance company, it becomes more important to study various attributes about their customers. Leveraging this customer information to make business decisions can provide a competitive edge to the company over other players in the market

We are provided with some customer data of an Insurance company like age, gender, BMI and medical charges billed by insurance company. We need to explore this data to see if we can derive some meaningful insights from this data.





## ALUM TALKS - Proof of Concept

Proof of concept

“For every Licensed user, our Company is charged for different software used by us. I manage the details of the number of licenses for different software being used in various accounts. After completing the SMDM course, I analyzed the cumulative costs, maximum and minimum costs of software licenses in different regions using various visualization techniques.

This made us realize that we were using 6 CRM software licenses where only 1 was sufficient at a particular location. I presented this to our management team and this saved us 4080 USD annually. I am really looking forward to upcoming modules in the Program. “

**Call for Action** - Please go back and think how you can use the concepts learned in this SMDM module, in your present role in your organization. - **Satish Kumar**



**ANY QUESTIONS**



**HAPPY LEARNING**