Qualitative

The information represents Characteristic.

Categorical data:

A countable number of distinct groups based on characterstics.

Binary data:-

Binary data can have only two values (1 & 0).

Ordinary data:-

It has three categories & The categories have a natural order.

Introduction to data visualization

Data visualization is basically putting The analyzed data in The form of visual like graphs, images. it easy for humans to understand The analyzed trends through visuals.

Benefits: Intuitive, Fast, Flexible

Plotting

Types of plots

Line plot, Scatter plot, ~~histoplot~~ histogram, boxplot, bar chart, pie chart.

Session-8

Simple scatter plots

Shows the relation ship between two variables.
it helps to compare data visualization.

Contour plots

Display Three dimensional data in two dimensions using contours (or) color coded regions.
Contour plots display the results of a single dataset (time step, load step, substep) over the model geometry.

. (all example programs in material)

Histograms

It is helped to make comparison in data sets over an interval (or) time. it helps to show a distribution of data

Binning

The intervals are also called bins. The bins are consective & non overlapping intervals of a variable. They must be adjacent & are often of equal size.

Density plot

It visualizes the distribution of data over a continious interval (or) time period. This chart is a variation of a Histogram that uses kernal smoothing to plot values.

Kernel Density Estimation (KDE)

It is a method for visualizing the distribution of observations in a dataset.

Session-9

Bar plot

The multiple numerical data columns can be compared against groups.

Comparision between Barplot & Histogram

→ Bar graph is pictorial representation of the data that uses bars to compare different categories of data.

→ Comparision of discrete variables.

→ categorical data

→ Bars do not touch each other

→ Elements are taken as individual entities.

→ Histogram refers to a graphical representation that displays the data by way of bars to show the frequency of numerical data.

→ Distribution of non discrete variables.

→ Quantitative data

→ Bars touch each other.

Boxplot

It provides a lot of information about any numerical data column. it is also known box and whisker plot. Minimum, first quartile, median, third quartile & maximum are extracted from the plots.

Jet colormap

-> View_colormap('jet')

-> Default in matplot prior to version 2.0,

Viridis colormap

-> Bright stripes in the grayscale image.

-> View_colormap ('Viridis')

Cubehelix & RdBu colormap

-> View_colormap ('cubehelix')

-> Showing +ve & -ve deviations from some mean, dual-color colorbars such as RdBu (Red blue) can be useful.

Session-10

Probability

$P(A)$ = Number of outcomes favorable to A / total number of possible outcomes $= \frac{n}{n+m}$

Statistical analysis for data

→ Statistical Analysis is the science of the exploration of the collection of large datasets to find different hidden patterns & trends.

Types of statistical Analysis

→ Quantitative Analysis.

→ Qualitative Analysis

Population (O) Sample data

→ if $B \subseteq A$ then $p(A) \geq p(B)$

→ $p(A \cup B) \leq p(A) + p(B)$

$P(A \cap B) \leq \min(P(A), P(B))$

$P(\neg A) = P(\Omega \setminus A) = 1 - P(A)$. is the set diff

Conditional probability

A conditional probability is the probability of an event occuring, given that another event has already occured.

Session 11

Bernoulli event

→ A Bernoulli event is one for which the probability the event occurs is p and the probability the event does not occur is $1-p$;

Bernoulli distribution

→ A Bernoulli distribution is the pair of probabilities of a Bernoulli event

$$P(z) = \begin{cases} p^z(1-p)^{1-z} & \text{for } z = 0, 1 \\ 0 & \text{otherwise.} \end{cases}$$

$E(x) = (1-p) \cdot 0 + p \cdot 1 = p$

→ For Bernoulli distribution $F(x)$ has

construction $F(x) = \begin{cases} 0 & \text{for } -\infty \leq x \leq 0 \\ 1-p & \text{for } 0 \leq x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$

Law of total probability

$$P(A) = P(A \cap B) + P(A \cap B')$$

~~Discussion Tree~~ Decision tree

→ The decision tree is a simple and convenient method of visualizing problems with the total probability rule.

Session-12

Random process

→ Random processes, also known as stochastic processes, are used to model uncertain quantities that evolve in time.

Random variable

→ A random variable x takes on a defined set of values with different probabilities.

Discrete random variables

→ The discrete random variables X and Y are said to be (stochastically) independent if, for all real numbers a & b

→ $P(\{X = a\} \cap \{Y = b\}) = P\{X = a\} \cdot P\{Y = b\}$

Normal distribution

→ A sample of outcomes from an experiment, a common first step is to plot the number of occurrences against sample values to get the distribution curve.

Poisson Distribution

$$P(X = k|\lambda) = f(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$