Nominal, ordinal, qualitative, and quantitative data are different types of data that are commonly used in statistics and data analysis.

1. Nominal data: Nominal data is a type of categorical data where the categories do not have any inherent order or ranking. Examples of nominal data include gender, eye color, or favorite color. Nominal data can be represented by labels or numbers, but the numbers themselves do not have any inherent meaning.
2. Ordinal data: Ordinal data is a type of categorical data where the categories have a natural order or ranking. Examples of ordinal data include rating scales, such as a 5-star rating system for a movie or restaurant. Ordinal data can be represented by labels or numbers, and the numbers do have a meaning in terms of their ranking.
3. Qualitative data: Qualitative data is descriptive data that is not easily measured or quantified. Examples of qualitative data include opinions, beliefs, and emotions. Qualitative data is often collected through open-ended questions or observations and is analyzed through coding and categorization.
4. Quantitative data: Quantitative data is numerical data that can be measured or quantified. Examples of quantitative data include height, weight, age, and income. Quantitative data can be further classified into discrete data, which can only take on specific values (e.g., the number of children in a family), and continuous data, which can take on any value within a range (e.g., temperature or time).

Understanding the different types of data is important in data analysis and statistics because different types of data require different methods of analysis and visualization.

Cross sectional data:
Data collected at the same time for many variables
eg:Amount of budget at particular time in 2022
Time series data:
data collected for single variable for different times
Different time points for eg:amount of rainfall for different years(2020 to 2022)
Weekly reports
Panel data:
Some are time independent and some are dependent for several variables.
Eg:unemlpoyment rate for several countries and several years.

Descriptive:
What happened to the data
Measure of central tendency(measure of location) : mean (simple mean, weighed mean (wixi/wi)if all weights are equal weighted mean converts to simple mean.and

trimmed mean(trimmed mean is mean after deleting outliers.)),
Geometric mean:x1.x2.....xn)power1/n
Harmonic mean:n/(1/x1+1/x2+...1/xn)
median (middle value)and mode(most occurring observation),
measure of shape, measure of spread

Disadvantage of mean:
Not for qualitative data
Affected by extreme values
If single observation is missing it will be difficult
Adv over median:
Doesn't affect if one value is missing or with outliers
Can't calculate with entire data only looks into middle value.

percentile(px)=x*(n+1)/100 where n is no of observations.
Decile:10 parts (10%,20%..)
Quartile:4 parts(25%,50%,75%,100%)

Mode is for qualitative data like gender

Mean-mode=3*(mean-median)

Data can be classified into:
Symmetric data, positively skewed and negatively skewed.
Symmetric data:mean, median and mode all lies on one point.
mean=median=mode

Positively skewed:
mean>median>mode

Negatively skewed:
Mean<median<mode

Measure of dispersion:

Range,iqr,variance and std

How other datapoints varies from central tendencies like mean
Location of other data points from mean
Range:max-min
Outliers —range is not good
Variance can be used with outliers

(Xi-xbar)pow2*(1/n-1) where bar is mean of sample

Coefficient of variation:
cv=(stddeviation/mean of sample)*100
Ration measure to spread
sigma=0 when no spread all values are equal otherwise >0
Sigma is for mean
Cv should be less for choosing
The coefficient of variation (CV) is a statistical measure that represents the relative variation of a dataset with respect to its mean. It is expressed as a percentage and is calculated as the ratio of the standard deviation (SD) to the mean, multiplied by 100:
CV = (SD / mean) x 100%
The CV is often used in fields such as finance, economics, and biology to compare the variability of different datasets, regardless of their scale or units of measurement. A low CV indicates that the data points are close to the mean, while a high CV indicates that the data points are more spread out from the mean.

Iqr:
Q3-q1
Outlier detection
Q1-1.5iqr<vaues<q3+1.5iqr
Outside outliers

Five number summary:
Q1,q2(median),q3,min,max

Box plot can be used to check five number summary

Kurtosis is a statistical measure that describes the degree of peakedness or flatness of a probability distribution relative to a normal distribution. A distribution with high kurtosis has a sharp peak and heavy tails, while a distribution with low kurtosis has a flatter peak and lighter tails.
The most commonly used measure of kurtosis is the fourth standardized moment, which is denoted by kurtosis or γ2. It is calculated as follows:
kurtosis = (1/n) Σ(xi - mean)^4 / (SD)^4 - 3
where xi represents each data point in the sample, n is the sample size, mean is the sample mean, and SD is the sample standard deviation.
A kurtosis of 3 indicates that the distribution has the same degree of peakedness as a normal distribution. A kurtosis greater than 3 indicates a distribution that is

more peaked than a normal distribution (i.e., leptokurtic), while a kurtosis less than 3 indicates a distribution that is less peaked than a normal distribution (i.e., platykurtic).

Platykurtic—thick tail—not to go with mean. K<3
Leptokurtic—thin tail k>3

nd=k=3

Probability:
Future outcome or prediction of likelihood of events
Model—data
Statistics:
Analysis of freq of past events
Data is given we have to find the model

Random variable is rule assigns numerical value for outcome of an interest
Discrete and continuous
Discrete:finite set of values
Classification dbms are discrete
Continuous:
Infinite set of values
Eg:time taken to resolve a ticket

For discrete rv:
Prob mass fn(p(X=xi)
Cumulative distribution fn
f(x)=p(X<=xi)
f(2)=p(0)+p(1)+p(2)

pdf(prob density fn for contieus rv):
Cumulative density fn is area covered with pdf
For continuous rv prob at particular variable is 0
f(x) or pdf >=0
p(a<=x<=b)=F(b)-F(a)
Integral f(x()dx=1

E(x)=xf(x)

var=(x-e(x))2

Prob mass fn:
Binomial (mean>variance),
multinominal,poisson

Poison:(mean=variance)
Outcome within time interval, region etc.
Eg:
No.of people visiting ticket counter in 8 hrs.

Properties:
No memory

Stats.poison.cdf(x,Lambda*t)

Continuous:
Pdf:
Continuous uniform distribution
f(x)=1/b-a if a<=x<=b
Otherwise 0
Mean is a+b/2 for uniform distribution
var=(b-a)2/12
Uniform.cdf(x=8,loc=0,scale=20

Normal dist:
Also Gaussian dist
Bell shaped curve
Eg:rainfall measurements

6sigma limit:(confidence intervals):

1sigma=68%
2sigma=95%
3sigma=99%

Total area above horizontal line under the curve=1

Std normal dist:
Z-transformation to calc prob(x1<x<x2) to avoid difficulty in normal dist for mean and variance
z=X-u/sigma
Std normal dist mean=0 var=1