STATISTICS WORKSHEET-4

1. What is the Central Limit Theorem?

The CLT is a statistical theory that states that - if you take a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from that population will be roughly equal to the population mean.

Consider there are 15 sections in class X, and each section has 50 students. Our task is to calculate the average marks of students in class X.

The standard approach will be to calculate the average simply:

- Calculate the total marks of all the students in Class X
- Add all the marks
- Divide the total marks by the total number of students

But what if the data is extremely large? Is this a good approach? No way, calculation marks of all the students will be a tedious and time-consuming process. So, what are the alternatives? Let's take a look at another approach.

- To begin, select groups of students from the class at random. This will be referred to as a sample. Create several samples, each with 30 students.
- Calculate each sample's individual mean.
- Calculate the average of these sample means.
- The value will give us the approximate average marks of the students in Class X.

- The histogram of the sample means marks of the students will resemble a bell curve or normal distribution

Important:-

The Central Limit Theorem is at the center of statistical inference what each data scientist/data analyst does every day**.**

The Central Limit Theorem is at the center of statistical inference what each data scientist/data analyst does every day

In this article, we will explore Central Limit Theorem, what is the Central Limit Theorem and why is it important and what is the difference between the Law of Large Numbers and the Central Limit Theorem?

The **Central Limit Theorem (CLT)** is a mainstay of **statistics** and **probability**. The theorem expresses that as the size of the sample expands, the distribution of the mean among multiple samples will be like a **Gaussian distribution**.

We can think of doing a trial and getting an outcome or an observation. We can rehash the test again and get another independent observation. Accumulated, numerous observations represent a sample of observations.

On the off chance that we calculate the mean of a sample, it will approximate the mean of the population distribution. In any case, like any estimate, it will not be right and will contain some mistakes. On the off chance that we draw numerous

independent samples, and compute their means, the distribution of those means will shape a Gaussian distribution.

The CLT gives us a certain distribution over our estimations. We can utilize this to pose an inquiry about the probability of an estimate that we make. For example, assume we are attempting to think about how an election will turn out.

question2. What is sampling? How many sampling methods do you know?

When you conduct research about a group of people, it's rarely possible to collect data from every person in that group. Instead, you select a **sample**. The sample is the group of individuals who will actually participate in the research.

To draw valid conclusions from your results, you have to carefully decide how you will select a sample that is representative of the group as a whole. This is called a **sampling method**. There are two primary types of sampling methods that you can use in your research:

- **Probability sampling** involves random selection, allowing you to make strong statistical inferences about the whole group.
- **Non-probability sampling** involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

question3. What is the difference between type1 and typeII error?

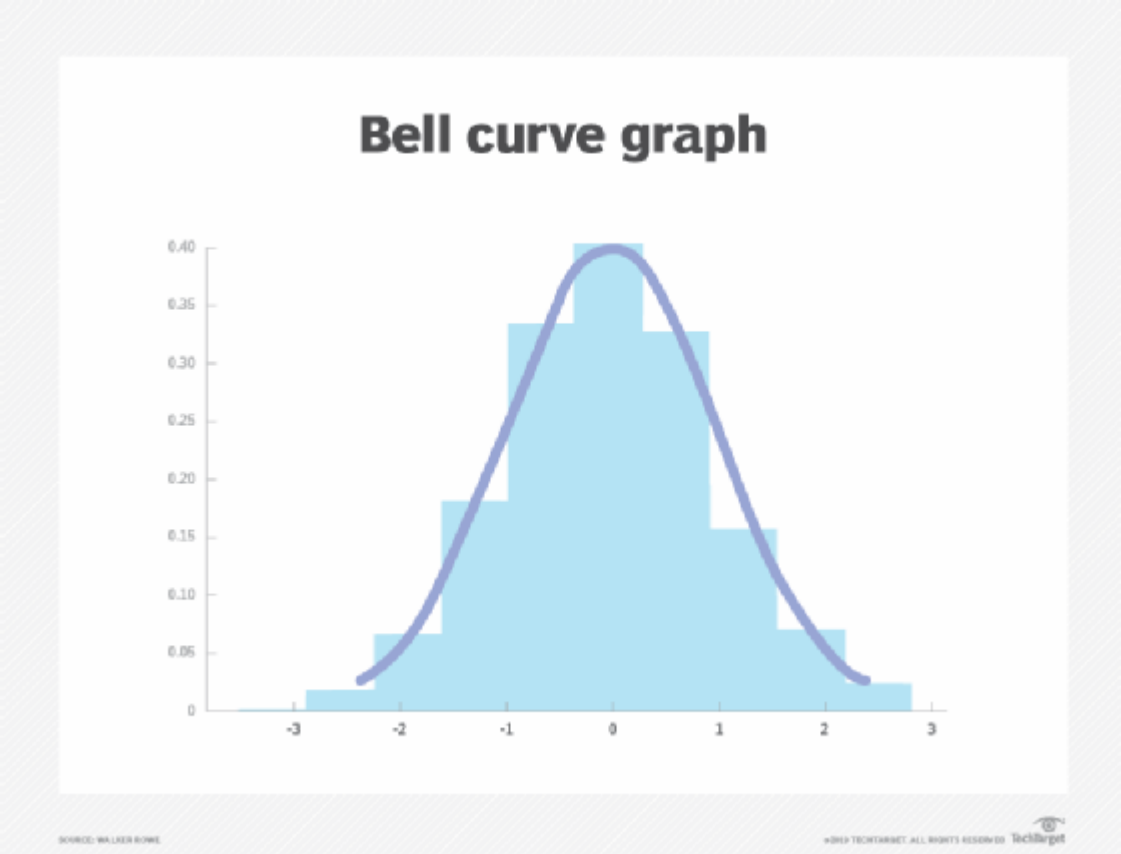
BASIS FOR COMPARISON	TYPE I ERROR	TYPE II ERROR

BASIS FOR COMPARISON	TYPE I ERROR	TYPE II ERROR
Meaning	Type I error refers to non-acceptance of hypothesis which ought to be accepted.	Type II error is the acceptance of hypothesis which ought to be rejected.
Equivalent to	False positive	False negative
What is it?	It is incorrect rejection of true null hypothesis.	It is incorrect acceptance of false null hypothesis.
Represents	A false hit	A miss
Probability of committing error	Equals the level of significance.	Equals the power of test.
Indicated by	Greek letter 'α'	Greek letter 'β'

question4. normal distribution?

A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the *mean* of the distribution.

The normal distribution is also known as a *Gaussian distribution* or *probability bell curve*. It is symmetric about the mean and indicates that values near the mean occur more frequently than the values that are farther away from the mean.



Graphically, a normal distribution is a bell curve because of its flared shape. The precise shape can vary according to the distribution of the values within the population. The population is the entire set of data points that are part of the distribution.

Regardless of its exact shape, a normal distribution bell curve is always symmetrical about the mean. A symmetrical distribution means that a vertical dividing line drawn through the maximum/mean value will produce two mirror images on either side of the line, in which half the population is less than the mean and half is greater. However, the reverse is not always true; that is, not all symmetrical distributions are normal. In the bell curve, the peak is always in the middle, and the mean, mode and median are all the same.

Normal distribution formula and empirical rule

The formula for the normal distribution is expressed below.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$ The formula for the normal distribution.

Here, *x* is value of the variable; *f(x)* represents the probability density function; *μ (mu)* is the mean; and *σ (sigma)* is the standard deviation.

The empirical rule for normal distributions describes where most of the data in a normal distribution will appear, and it states the following:

- 68.2% of the observations will appear within +/-1 standard deviation of the mean;
- 95.4% of the observations will fall within +/-2 standard deviations; and
- 99.7% of the observations will fall within +/-3 standard deviations.

All data points falling outside of three standard deviations (3σ) indicate rare occurrences

question5. **Covariance** :-Covariance measures how the two variables move concerning each other and is an extension of the concept of variance (which tells about how a single variable varies). It can take any value from -∞ to +∞.

- The higher this value, the more dependent the relationship is. A positive number signifies positive covariance and denotes a direct connection. Effectively this means that an increase in one

variable would also lead to a corresponding increase in the other variable, provided other conditions remain constant.

- On the other hand, a negative number signifies negative covariance, which denotes an inverse relationship between the two variables. Though covariance is perfect for defining the type of relationship, it is not good for interpreting its magnitude.

  Correlation :-Correlation is a step ahead of covariance as it quantifies the relationship between two random variables. In simple terms, it is a unit measure of how these variables change concerning each other (normalized covariance value).

  - The correlation has an upper and lower cap on a range, unlike covariance. It can only take values between +1 and -1. A correlation of +1 indicates that random variables have a direct and strong relationship.
  - On the other hand, the correlation of -1 indicates a strong inverse relationship, and an increase in one variable will lead to an equal and opposite decrease in the other variable. 0 means that the two numbers are independent.

- The Formula for Covariance and Correlation

  Let us express these concepts mathematically for two random variables, A and B, with mean values as Ua and Ub and standard deviation as Sa and Sb, respectively.

  Effectively we can define the relationship between the two:

  $$\text{Covariance} = \text{Correleation} * S_a * S_b$$

  Both correlations and covariance find application in statistical and financial analysis fields. Since correlation standardizes the connection, it is helpful in comparison of any two variables. In addition, it helps analysts develop strategies like pair trade and **hedging** for efficient returns on the portfolio and safeguarding these returns in terms of adverse movements in the stock market

question**6. . Differentiate between univariate ,Biavariate,and multivariate analysis**

This type of data consists of **only one variable**. The analysis of univariate data is thus the simplest form of analysis since the information deals with only one quantity that changes. It does not deal with causes or relationships and the main purpose of the analysis is to describe the data and find patterns that exist within it. The example of a univariate data can be height.

Heights (in cm)	164	167.3	170	174.2	178	180	186

Suppose that the heights of seven students of a class is recorded(figure 1),there is only one variable that is height and it is not dealing with any cause or relationship. The description of patterns found in this type of data can be made by drawing conclusions using central tendency measures (mean, median and mode), dispersion or spread of data (range, minimum, maximum, quartiles, variance and standard deviation) and by using frequency distribution tables, histograms, pie charts, frequency polygon and bar charts.

2. Bivariate data –
This type of data involves **two different variables**. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables.Example of bivariate data can be temperature and ice cream sales in summer season.

TEMPERATURE(IN CELSIUS)	ICE CREAM SALES
20	2000
25	2500
35	5000
43	7800

Suppose the temperature and ice cream sales are the two variables of a bivariate data(figure 2). Here, the relationship is visible from the table that temperature and sales are directly proportional to each other and thus related because as the temperature increases, the sales also increase. Thus bivariate data analysis involves comparisons, relationships, causes and explanations. These variables are often plotted on X and Y axis on the graph for better understanding of data and one of these variables is independent while the other is dependent.

3. Multivariate data –
When the data involves **three or more variables**, it is categorized under multivariate. Example of this type of data is suppose an advertiser wants to compare the popularity of four

advertisements on a website, then their click rates could be measured for both men and women and relationships between variables can then be examined.
It is similar to bivariate but contains more than one dependent variable. The ways to perform analysis on this data depends on the goals to be achieved.Some of the techniques are regression analysis,path analysis,factor analysis and multivariate analysis of variance (MANOVA).

question7. The technique used to determine how independent variable values will impact a particular dependent variable under a given set of assumptions is defined as **sensitive analysis**. It's usage will depend on one or more input variables within the specific boundaries, such as the effect that changes in interest rates will have on a bond's price.

It is also known as the what – if analysis. Sensitivity analysis can be used for any activity or system. All from planning a family vacation with the variables in mind to the decisions at corporate levels can be done through sensitivity analysis.

Assuming this function is is n= f(A, lmbda) , Then the amplitude sensitivity S at a given **lambda= (dn/n) /(dA/A)**, one can calculate S at different lambdas and plot S versus lambda. This is the amplitude sensitivity lambda curve

Methods of Sensitivity Analysis

There are different methods to carry out the sensitivity analysis:

- Modeling and simulation techniques
- Scenario management tools through Microsoft excel

There are mainly two approaches to analyzing sensitivity:

- Local Sensitivity Analysis
- Global Sensitivity Analysis

Local sensitivity analysis is derivative based (numerical or analytical). The term local indicates that the derivatives are

taken at a single point. This method is apt for simple cost functions, but not feasible for complex models, like models with discontinuities do not always have derivatives.

Mathematically, the sensitivity of the cost function with respect to certain parameters is equal to the partial derivative of the cost function with respect to those parameters.

Local sensitivity analysis is a *one-at-a-time* (OAT) technique that analyzes the impact of one parameter on the cost function at a time, keeping the other parameters fixed.

Global sensitivity analysis is the second approach to sensitivity analysis, often implemented using Monte Carlo techniques. This approach uses a global set of samples to explore the design space.

The various techniques widely applied include:

- ***Differential sensitivity analysis:*** It is also referred to the direct method. It involves solving simple partial derivatives to temporal sensitivity analysis. Although this method is computationally efficient, solving equations is intensive task to handle.
- ***One at a time sensitivity measures:*** It is the most fundamental method with partial differentiation, in which varying parameters values are taken one at a time. It is also called as local analysis as it is an indicator only for the addressed point estimates and not the entire distribution.
- ***Factorial Analysis:*** It involves the selection of given number of samples for a specific parameter and then running the model for the combinations. The outcome is then used to carry out parameter sensitivity.

Calculation of the Sensitivity Analysis (Step by Step)

Lets start.

1. **Firstly, the analyst is required to design the basic formula, which will act as the output formula. For instance, say NPV formula can be taken as the output formula.**
2. **Next, the analyst needs to identify which are the variables that are required to be sensitized as they are key to the output formula. In the NPV formula in excel, the cost of capital and the initial investment can be the independent variables.**

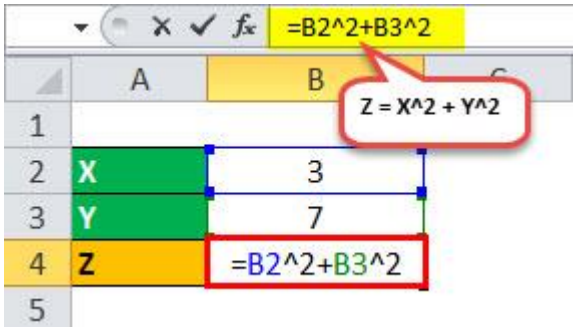
3. **Next, determine the probable range of the independent variables.**
4. **Next, open an excel sheet and then put the range of one of the independent variable along the rows and the other set along with the columns.**

- Range of 1st independent variable
- Range of 2nd independent variable

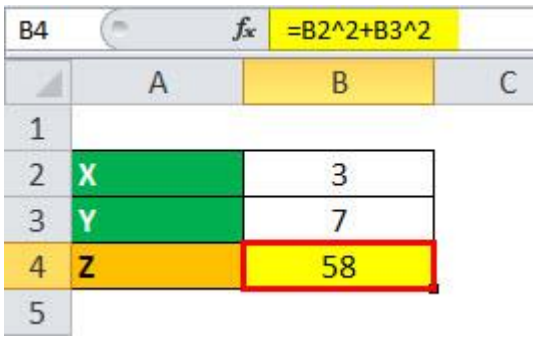
Example #1

Let us take the example of a simple output formula, which is stated as the summation of the square of two independent variables X and Y.

In this case, let us assume the range of X as 2, 4, 6, 8, and 10, while that of Y as 1, 3, 5, 7, 9, 11, and 13. Based on the above-mentioned technique, all the combinations of the two independent variables will be calculated to assess the sensitivity of the output.



For instance, if X = 3 (Cell B2) and Y = 7 (Cell B3), then Z = $3^2 + 7^2 = 58$ (Cell B4)



Z = 58

For the calculation of Sensitivity Analysis, go to the Data tab in excel and then select What if analysis option. For the further procedure of sensitivity analysis calculation, refer to the given article here – **Two-Variable Data Table in Excel**

	A	B	C	D	E	F	G	H
7			X					
8		58	2	4	6	8	10	
9	Y	1	5	17	37	65	101	
10		3	13	25	45	73	109	
11		5	29	41	61	89	125	
12		7	53	65	85	113	149	
13		9	85	97	117	145	181	
14		11	125	137	157	185	221	
15		13	173	185	205	233	269	
16								

question8. hypothesis testing :=-A statistical hypothesis is an assertion or conjecture concerning one or more populations. To prove that a hypothesis is true, or false, with absolute certainty, we would need absolute knowledge. That is, we would have to examine the entire population. Instead, hypothesis testing concerns on how to use a random sample to judge if it is evidence that supports or not the hypothesis

Hypothesis testing is formulated in terms of two hypotheses:

• H0: the null hypothesis;

• H1: the alternate hypothesis.

The hypothesis we want to test is if H1 is “likely” true. So, there are two possible outcomes:

• Reject H0 and accept H1 because of sufficient evidence in the sample in favor or H1;

• Do not reject H0 because of insufficient evidence to support H1

Very important!! Note that failure to reject H0 does not mean the null hypothesis is true. There is no formal outcome that says “accept H0.” It only means that we do not have sufficient evidence to support H1.

Because we are making a decision based on a finite sample, there is a possibility that we will make mistakes. The possible outcomes are: H0 is true H.

Two-tailed test: example (cont.)

We know the sampling distribution of $\bar{X}$ is a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n} = 0.8$ due to the central limit theorem.

Therefore we can compute the probability of a type I error as α = Pr($\bar{X}$ 102 when μ = 100)

= Pr Z < 98 − 100 8/ √ 100) + Pr(Z > 102 − 100 8/ √ 100 □

= Pr(Z < −2.5) + Pr(Z > 2.5)

= 2 × Pr(Z < −2.5) = 2 × 0.0062 = 0.0124.

question 9. Qualitative data :-Qualitative data is non-statistical and is typically unstructured or semi-structured. This data isn’t necessarily measured using hard numbers used to develop graphs and charts. Instead, it is categorized based on properties, attributes, labels, and other identifiers.

Qualitative data can be used to ask the question “why.” It is investigative and is often open-ended until further research is conducted. Generating this data from qualitative research is used for theorizations, interpretations, developing hypotheses, and initial understandings.

Qualitative data can be generated through:

- Texts and documents
- Audio and video recordings
- Interview transcripts and focus groups
- Observations and notes

Qualitative data examples

To better understand qualitative data, let's take the example of a bookcase. The following characteristics of this bookcase determine the quality of the information that's available to us about it:

- Made of wood
- Built in Italy
- Deep brown
- Golden knobs
- Smooth finish
- Made of oak

quantitative data

Contrary to qualitative data, quantitative data is statistical and is typically structured in nature – meaning it is more rigid and defined. This data type is measured using numbers and values, making it a more suitable candidate for data analysis.

Whereas qualitative is open for exploration, quantitative data is much more concise and close-ended. It can be used to ask the questions “how much” or “how many,” followed by conclusive information

Quantitative data can be generated through:

- Tests
- Experiments
- Surveys
- Market reports
- Metrics

Types of quantitative data and examples

Quantitative data can be broken into further sub-categories. These categories are called discrete and continuous data.

Discrete data

Discrete data is just data that cannot be broken down into smaller parts. This type of data consists of integers (positive and negative numbers, e.g., -100, 10, 100, and so on) and is finite (meaning it reaches a limit).

A few examples of discrete data would be how much change you have in your pocket, how many iPhones were sold last year, and how much traffic came to your website today.

Another important note is that discrete data can technically be categorical. For example, the number of baseball players in a team born in Mexico is whole and discrete.

Continuous data

Continuous data is data that can be infinitely broken down into smaller parts or data that continuously fluctuates.

A few examples of continuous data would be the speed of your train during the morning commute, the time you take to write an article, your weight, and your age.

question 10. How to calculate range and interquartile range?

The **quartiles** of a ranked set of data values are three points which divide the data into exactly four equal parts, each part comprising of quarter data.

1. **Q1** is defined as the middle number between the smallest number and the median of the data set.
2. **Q2** is the median of the data.
3. **Q3** is the middle value between the median and the highest value of the data set.
   The Python **range() function** returns a sequence of numbers, in a given range. The most common use of it is to iterate sequence on a sequence of numbers using Python loops.

Syntax: *range(start, stop, step)*
Parameter:

- ***start****: [optional] start value of the sequence*
- ***stop:*** *next value after the end value of the sequence*
- ***step:*** *[optional] integer value, denoting the difference between any two numbers in the sequence.*

Return: *Returns a range type object.*

```
# print first 5 integers
# using python range() function
for i in range(5):
    print(i, end=" ")
print()
```

The **interquartile range** The **interquartile range**, often denoted “IQR”, is a way to measure the spread of the middle 50% of a dataset. It is calculated as the difference between the

first quartile* (the 25th percentile) and the third quartile (the 75th percentile) of a dataset.

Fortunately it's easy to calculate the interquartile range of a dataset in Python using the numpy.percentile() function.

This tutorial shows several examples of how to use this function in practice.

```
import numpy as np

#define array of data
data = np.array([14, 19, 20, 22, 24, 26, 27, 30, 30, 31, 36, 38, 44, 47])

#calculate interquartile range
q3, q1 = np.percentile(data, [75 ,25])
iqr = q3 - q1

#display interquartile range
iqr

12.25
```

question 11.Bell Curve?

A bell curve is a common type of distribution for a variable, also known as the normal distribution. The term "bell curve" originates from the fact that the graph used to depict a normal distribution consists of a symmetrical bell-shaped curve.

The highest point on the curve, or the top of the bell, represents the most probable event in a series of data (its mean, mode, and median in this case), while all other possible occurrences are symmetrically distributed around the mean, creating a downward-sloping curve on each side of the

peak. The width of the bell curve is described by its standard deviation.

KEY TAKEAWAYS

- A bell curve is a graph depicting the normal distribution, which has a shape reminiscent of a bell.
- The top of the curve shows the mean, mode, and median of the data collected.
- Its standard deviation depicts the bell curve's relative width around the mean.
- Bell curves (normal distributions) are used commonly in statistics, including in analyzing economic and financial data.

Understanding a Bell Curve

The term "bell curve" is used to describe a graphical depiction of a normal probability distribution, whose underlying standard deviations from the mean create the curved bell shape. A standard deviation is a measurement used to quantify the variability of data dispersion, in a set of given values around the mean. The mean, in turn, refers to the average of all data points in the data set or sequence and will be found at the highest point on the bell curve

Example of a Bell Curve

A bell curve's width is defined by its standard deviation, which is calculated as the level of variation of data in a sample around the mean. Using the empirical rule, for example, if 100 test scores are collected and used in a normal probability distribution, 68% of those test scores should fall within one standard deviation above or below the mean. Moving two standard deviations away from the mean should include 95% of the 100 test scores collected. Moving three standard deviations away from the mean should represent 99.7% of the scores (see the figure above).

Test scores that are extreme outliers, such as a score of 100 or 0, would be considered long-tail data points that consequently lie squarely outside of the three standard deviation range.

question 12. Outliers are values at the extreme ends of a dataset.

Some outliers represent true values from natural variation in the population. Other outliers may result from incorrect data entry, equipment malfunctions, or other measurement errors.

An outlier isn't always a form of dirty or incorrect data, so you have to be careful with them in data cleansing. What you should do with an outlier depends on its most likely cause.

True outliers

True outliers should always be retained in your dataset because these just represent natural variations in your sample.

Four ways of calculating outliers

You can choose from several methods to detect outliers depending on your time and resources.

Sorting method

You can **sort** quantitative variables from low to high and scan for extremely low or extremely high values. Flag any extreme values that you find.

This is a simple way to check whether you need to investigate certain data points before using more sophisticated methods.

Using visualizations

You can use software to **visualize** your data with a box plot, or a box-and-whisker plot, so you can see the data distribution at a glance. This type of chart highlights minimum and maximum values (the range), the median, and the interquartile range for your data.

Many computer programs highlight an outlier on a chart with an asterisk, and these will lie outside the bounds of the graph.

Statistical outlier detection

Statistical outlier detection involves applying **statistical tests** or procedures to identify extreme values.

You can convert extreme data points into *z* scores that tell you how many standard deviations away they are from the mean.

If a value has a high enough or low enough *z* score, it can be considered an outlier. As a rule of thumb, values with a z score greater than 3 or less than –3 are often determined to be outliers.

Using the interquartile range

The **interquartile range** (IQR) tells you the range of the middle half of your dataset. You can use the IQR to create “fences” around your data and then define outliers as any values that fall outside those fences.

question 13. p-value in Python Statistics
When talking statistics, a p-value for a statistical model is the probability that when the null hypothesis is true, the statistical summary is equal to or greater than the actual observed results. This is also termed ‘*probability value*’ or ‘*asymptotic significance*’.

All statistical tests have a null hypothesis. For most tests, the null hypothesis is that there is no relationship between your variables of interest or that there is no difference among groups.

For example, in a two-tailed *t* test, the null hypothesis is that the difference between two groups is zero.

xample: Null and alternative hypothesisYou want to know whether there is a difference in longevity between two groups of mice fed on different diets, diet A and diet B. You can statistically test the difference between these two diets using a two-tailed t test.

- **Null hypothesis (H_0):** there is no difference in longevity between the two groups.
- **Alternative hypothesis (H_A or H_1):** there is a difference in longevity between the two groups.

Caution when using *p* values

P values are often interpreted as your risk of rejecting the null hypothesis of your test when the null hypothesis is actually true.

In reality, the risk of rejecting the null hypothesis is often higher than the *p* value, especially when looking at a single study or when using small sample sizes. This is because the smaller your frame of reference, the greater the chance that you stumble across a statistically significant pattern completely by accident.

P values are also often interpreted as supporting or refuting the alternative hypothesis. This is not the case. **The *p* value can only tell you whether or not the null hypothesis is supported.** It cannot tell you whether your alternative hypothesis is true, or why.

Reporting *p* values

P values of statistical tests are usually reported in the results section of a research paper, along with the key information needed for readers to put the *p* values in context – for example, correlation coefficient in a linear regression, or the average difference between treatment groups in a *t*-test.

Example: Reporting the resultsIn our comparison of mouse diet A and mouse diet B, we found that the lifespan on diet A (M = 2.1 years; SD = 0.12) was significantly shorter than the lifespan

on diet B (M = 2.6 years; SD = 0.1), with an average difference of 6 months ($t(80) = -12.75$; $p < 0.01$).

question 14.

Binomial Probability

Binomial probability refers to the probability of exactly xx successes on nn repeated trials in an experiment which has two possible outcomes (commonly called a binomial experiment).
If the probability of success on an individual trial is pp , then the binomial probability is nCx·px·(1−p)n−xnCx·px·(1−p)n−x .
Here nCxnCx indicates the number of different combinations of xx objects selected from a set of nn objects. Some textbooks use the notation (nx)(nx) instead of nCxnCx .
Note that if pp is the probability of success of a single trial, then (1−p)(1−p) is the probability of failure of a single trial.

What is the probability of getting 66 heads, when you toss a coin 1010 times?
In a coin-toss experiment, there are two outcomes: heads and tails. Assuming the coin is fair , the probability of getting a head is 1212 or 0.50.5 .
The number of repeated trials: n=10n=10
The number of success trials: x=6x=6
The probability of success on individual trial: p=0.5p=0.5
Use the formula for binomial probability.

10C6·(0.5)6·(1−0.5)10−610C6·(0.5)6·(1−0.5)10−6
Simplify.

≈0.205
If the outcomes of the experiment are more than two, but can be broken into two probabilities pp and qq such

that $p+q=1$, the probability of an event can be expressed as binomial probability.

For example, if a six-sided die is rolled 10 times, the binomial probability formula gives the probability of rolling a three on 4 trials and others on the remaining trials.
The experiment has six outcomes. But the probability of rolling a 3 on a single trial is $\frac{1}{6}$ and rolling other than 3 is $\frac{5}{6}$.
Here, $\frac{1}{6}+\frac{5}{6}=1$.
The binomial probability is:

$$10C4\cdot(\frac{1}{6})^4\cdot(1-\frac{1}{6})^{10-4}$$
Simplify.

$$\approx 0.05$$

question 15. What Is Analysis of Variance (ANOVA)?

Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

The t- and z-test methods developed in the 20th century were used for statistical analysis until 1918, when Ronald Fisher created the analysis of variance method.[1][2] ANOVA is also called the Fisher analysis of variance, and it is the extension of the t- and z-tests. The term became well-known in 1925, after appearing in Fisher's book, "Statistical Methods for Research Workers."[3] It was employed in experimental psychology and later expanded to subjects that were more complex.

What Is the Analysis of Variance (ANOVA)?

The Formula for ANOVA is:

$$F=MSE/MST$$

where:

F=ANOVA coefficient

MST=Mean sum of squares due to treatment

MSE=Mean sum of squares due to error

Measurement of sensitivity analysis

Below are mentioned the steps used to conduct sensitivity analysis:

1. Firstly the base case output is defined; say the NPV at a particular base case input value (V1) for which the sensitivity is to be measured. All the other inputs of the model are kept constant.
2. Then the value of the output at a new value of the input (V2) while keeping other inputs constant is calculated.
3. Find the percentage change in the output and the percentage change in the input.
4. The sensitivity is calculated by dividing the percentage change in output by the percentage change in input.