**ASSIGNMENT 1 - Applying Statistics using Scipy**

**Questions To Add your Responses :**

1. Explain what is the Central Limit Theorem to a non-technical audience?

1. Central Limit Theorem (CLT)

Explanation: The Central Limit Theorem (CLT) states that the distribution of sample means will approximate a normal distribution, regardless of the population's distribution, as long as the sample size is sufficiently large (usually n > 30).

Example: Suppose you want to know the average height of adult men in a city. The population distribution of heights might be skewed. If you take random samples of 50 men and calculate their average heights, and repeat this process many times, the distribution of those average heights will form a normal (bell-shaped) curve.

## **2. What is Sampling and following techniques required to gather sample data?**

Sampling is the process of selecting a subset of individuals or observations from a larger population to estimate characteristics of the whole population.

#### **Sampling Techniques:**

* **Simple Random Sampling**: Every member of the population has an equal chance of being selected. E.g., drawing names from a hat.
* **Stratified Sampling**: The population is divided into subgroups (strata) based on a characteristic, and samples are drawn from each subgroup. E.g., sampling students from different grade levels.
* **Systematic Sampling**: Every nth member of the population is selected. E.g., selecting every 10th person on a list.
* **Cluster Sampling**: The population is divided into clusters, some clusters are randomly selected, and all members of selected clusters are sampled. E.g., surveying randomly selected classrooms in a school.

3. Describe the Hypothesis Testing and why do we conduct it?

Hypothesis testing is a statistical method used to make decisions or inferences about a population based on sample data. It involves the following steps:

* **Formulate Hypotheses**: The null hypothesis (H0) represents no effect or no difference, while the alternative hypothesis (H1) represents the effect or difference.
* **Select Significance Level**: Usually set at 0.05, indicating a 5% risk of concluding that an effect exists when it doesn't.
* **Calculate Test Statistic**: Based on the sample data.
* **Determine p-value**: The probability of observing the data if the null hypothesis is true.
* **Compare p-value with Significance Level**: If p-value < significance level, reject H0; otherwise, do not reject H0.

**Example**:

* **Null Hypothesis (H0)**: The average income in a city is $50,000.
* **Alternative Hypothesis (H1)**: The average income in a city is not $50,000.
* **Procedure**: Collect a sample of incomes, calculate the sample mean, and perform a statistical test (e.g., t-test).
* **Outcome**: If the p-value is less than 0.05, reject H0; otherwise, do not reject H0.

Hypothesis testing helps determine whether there is enough evidence to support a specific claim about a population.

4. Define and briefly elaborate Central Tendency using measures with examples?

**Measures of Central Tendency**: These measures summarize a data set with a single value representing the center.

* **Mean**:
  + **Example**: The mean score of 5 students' test scores (80, 85, 90, 95, 100) is 80+85+90+95+1005=90\frac{80+85+90+95+100}{5} = 90580+85+90+95+100​=90.
* **Median**:
  + **Example**: The median of the test scores (80, 85, 90, 95, 100) is 90, the middle value.
* **Mode**:
  + **Example**: In a set of shoe sizes (6, 7, 7, 8, 9), the mode is 7, as it appears most frequently.

5. What’s a ROC Curve? How do you differentiate Point Estimate and Confidence Interval Estimate?

**ROC Curve**: A graphical plot that shows the performance of a binary classification model by plotting the true positive rate against the false positive rate at various threshold settings.

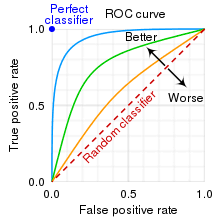
**Example**: Evaluating a medical test for diagnosing a disease, the ROC curve helps visualize the trade-off between sensitivity and specificity.

**Point Estimate**: A single value estimate of a population parameter.

* **Example**: The sample mean height of 100 men is a point estimate of the population mean height.

**Confidence Interval Estimate**: A range of values that likely contains the population parameter.

* **Example**: If the sample mean height of 100 men is 175 cm with a standard error of 2 cm, a 95% confidence interval might be 175±1.96×2175 \pm 1.96 \times 2175±1.96×2 cm, or (171.08, 178.92) cm.



6. What’s the difference between Correlation and Covariance in Statistics? What are the goals of A/B Testing?

**Correlation**: Measures the strength and direction of a linear relationship between two variables, standardized between -1 and 1.

* **Example**: Correlation between hours studied and exam scores is 0.85, indicating a strong positive relationship.

**Covariance**: Measures how two variables change together, not standardized.

* **Example**: Covariance between height and weight might be 100, indicating a positive relationship but not providing a standard measure.

**A/B Testing Goals**:

* **Goal**: Determine which version (A or B) of a variable (like a webpage) performs better.
* **Example**: Test two different landing pages (A and B) to see which has a higher conversion rate.

7. How would you build and test a metric to compare two user’s ranked lists of movie/tv show preferences?

To compare two users' ranked lists of movie/TV show preferences, you can use metrics like **Kendall Tau** or **Spearman’s Rank Correlation**.

#### **Steps:**

1. **Obtain Rankings**: Get ranked lists from both users.
2. **Calculate Metric**: Use a formula to calculate the similarity between the two lists.
3. **Interpret Results**: Higher similarity scores indicate more agreement between the lists.

To compare two users' ranked lists of movie/TV show preferences, you can use **Spearman’s Rank Correlation**.

**Example**:

* **User A Rankings**: [1, 2, 3, 4, 5]
* **User B Rankings**: [2, 1, 4, 5, 3]

Calculate Spearman's rank correlation coefficient (ρ) using the formula:

ρ=1−6∑di2n(n2−1)\rho = 1 - \frac{6 \sum d\_i^2}{n(n^2 - 1)}ρ=1−n(n2−1)6∑di2​​

where did\_idi​ is the difference between ranks of each item, and nnn is the number of items.

8. Find the probability of P(x<400) given that mean is = 1000 variance is =100

Given a normal distribution with mean (μ\muμ) = 1000 and variance (σ2\sigma^2σ2) = 100, find P(X<400)P(X < 400)P(X<400).

1. **Standardize the Variable**: Convert to a standard normal distribution: Z=X−μσ=400−1000100=400−100010=−60Z = \frac{X - \mu}{\sigma} = \frac{400 - 1000}{\sqrt{100}} = \frac{400 - 1000}{10} = -60Z=σX−μ​=100​400−1000​=10400−1000​=−60
2. **Find Probability**: Using Z-tables, P(Z<−60)P(Z < -60)P(Z<−60) is virtually 0, as -60 is far in the tail of the normal distribution.

9. Provide a proper description as to when to apply Ztest, t-test, Chi-Square and annova, with examples.

**Z-test: Used for large samples (n > 30) or known population variance.**

* **Example:** Testing if the average height of a sample of 100 people differs from the known average height of 170 cm.

**t-test: Used for small samples (n ≤ 30) and unknown population variance.**

* **Example:** Comparing the average test scores of 20 students in two different classes.

**Chi-Square Test:** Used for categorical data to test independence or goodness-of-fit.

* **Example:** Testing if gender is related to voting preference in an election.

**ANOVA (Analysis of Variance): Used to compare means of three or more groups.**

* **Example:** Comparing the average test scores of students from three different schools.

**Explain the questions and problems in your approach while solving them. No copy paste allowed. Description of the outcome is really important to get an understanding better.**

**Assignment 1**

**SET A**

**Strings :-**

**Q1) Write a python program to check whether the string is Symmetrical or Palindrome**

**Q2) Write a Python program to reverse the words in a given string**

**Q3) Write a Python program to remove the i’th character from the string in different ways**

**Functions :-**

**Q1) Write a python function to find the Max of three numbers**

**Q2) Write a python function to Sum all the numbers in a list**

**Q3) Write a Python function to reverse the string**

**SET B**

**Strings :--**

**Q1) Write a python program to print even length words in a string**

**Q2) Write a python program to accept the strings which contains all vowels**

**Q3) Write a python program to Count the number of matching characters in a pair of String**

**Functions :--**

**Q1) Write a Python function that takes a list and returns a new list with unique elements of the first list.**

**Q2) Write a python function that takes a number as a parameter and check the number is prime or not.**

**Q3) Write a python function to check whether a number is perfect or not.**

**Assignment 2**

**SET A 1. Python Program to Calculate the Area of a Triangle**

**2. Python Program to Swap Two Variables**

**3. Python Program to Generate a Random Number**

**SET B**

**1. Write a Python Program to Check if a Number is Positive, Negative or Zero**

**2. Write a Python Program to Check if a Number is Odd or Even**

**3. Write a Python Program to Check Prime Number**

**4. Write a Python Program to Check Armstrong Number**

**5. Write a Python Program to Find the Factorial of a Number**

**PROGRAMS FOR PRACTICE:**

**1. Python Program to Convert Kilometers to Miles**

**2. Python Program to Convert Celsius To Fahrenheit**

**3. Write a Python Program to Check Leap Year**

**4. Write a Python Program to Print all Prime Numbers in an Interval**

**5. Write a Python Program to Print the Fibonacci sequence**

**6. Write a Python Program to Find Armstrong Number in an Interval**

**7. Write a Python Program to Find the Sum of Natural Numbers**

**Assignment 3**

**1. Write a Python program to append items from a specified list.**

**2. Write a python program Check if a Substring is Present in a Given String**

**3. Write a python program Words Frequency in String Shorthands**

**4. Write a python program Convert Snake case to Pascal case**

**5. Write a Python function to calculate the factorial of a number (a non-negative integer). The function accepts the number as an argument.**

**6. Write a Python function to check whether a number is in a given range.**

**7. Write a Python function that accepts a string and calculate the number of upper case letters and lower case letters.**

**8. Write a Python program to detect the number of local variables declared in a function.**

**8. Write a python program to Remove all duplicates from a given string in Python**

**9. Write a Python function that checks whether a passed string is palindrome or not.**

**11. Write a Python program that accepts a hyphen-separated sequence of words as input and prints the words in a hyphen-separated sequence after sorting them alphabetically.**

**Assignment 4**

**SET A**

**List**

**1) Write a Python program to sum all the items in a list.**

**2) Write a Python program to multiplies all the items in a list.**

**3) Write a Python program to get a list, sorted in increasing order by the last element in each tuple from a**

**given list of non-empty tuples.**

**Tuples:--**

**1) Write a Python program to create a tuple.**

**2) Write a Python program to create a tuple with different data types.**

**3) Write a Python program to check whether an element exists within a tuple.**

**Sets:--**

**1) Write a Python program to create a set.**

**2) Write a Python program to iterate over sets.**

**4) Write a Python program to create set difference.**

**Dictionary:--**

**1) Write a Python script to sort (ascending and descending) a dictionary by value.**

**2) Write a Python script to add a key to a dictionary.**

**3) Write a Python program to iterate over dictionaries using for loops.**

**SET B**

**List**

**1. Write a Python program to remove duplicates from a list.**

**2. Write a Python program to check a list is empty or not.**

**Tuples**

**1. Write a Python program to convert a list to a tuple.**

**2. Write a Python program to remove an item from a tuple.**

**3. Write a Python program to slice a tuple.**

**4. Write a Python program to find the length of a tuple.**

**Sets**

**1. Write a Python program to check if a set is a subset of another set.**

**2. Write a Python program to find maximum and the minimum value in a set.**

**3. Write a Python program to find the length of a set.**

**Dictionary**

**1. Write a Python script to generate and print a dictionary that contains a number (between 1 and n) in the form**

**(x, x\*x).**

**3. Write a Python script to merge two Python dictionaries.**

**4. Write a Python program to get a dictionary from an object's fields.**

**PROGRAMS FOR PRACTICE:**

**1. Write a Python program to get the largest number from a list.**

**2. Write a Python program to get the smallest number from a list.**

**3. Write a Python program to count the number of strings where the string length is 2 or more and the**

**first and last character are same from a given list of strings.**

**3. Write a Python program to add an item in a tuple.**

**4. Write a Python program to convert a tuple to a string.**

**5. Write a Python program to create the colon of a tuple.**

**6. Write a Python program to unpack a tuple in several variables.**

**7. Write a Python program to add member(s) in a set.**

**8. Write a Python program to remove item(s) from set**

**9. Write a Python program to create an intersection of sets.**

**11. Write a Python program to create a union of sets.**

**12. Write a Python script to concatenate following dictionaries to create a new one.**

**13. Write a Python program to map two lists into a dictionary.**

**14. Write a Python program to sort a dictionary by key.**

**15. Write a Python program to get the maximum and minimum value in a dictionary.**

**16. Write a Python program to clone or copy a list.**

**17. Write a Python program to find the list of words that are longer than n from a given list of words.**

**18. Write a Python program to unzip a list of tuples into individual lists.**

**19. Write a Python program to reverse a tuple.**

**20. Write a Python program to convert a list of tuples into a dictionary.**

**21. Write a Python program to print a tuple with string formatting.**

**22. Write a Python program to create a symmetric difference.**

**23. Write a Python program to check if a given value is present in a set or not.**

**24. Write a Python program to check if a given set is superset of itself and superset of another given set.**

**25. Write a Python program to check a given set has no elements in common with other given set.**

**26. Write a Python program to remove the intersection of a 2nd set from the 1st set.**

**27. Write a Python program to remove duplicates from Dictionary.**

**28. Write a Python script to check whether a given key already exists in a dictionary.**

**29. Write a Python program to sum all the items in a dictionary.**

**30. Write a Python program to multiply all the items in a dictionary.**

**31. Write a Python program to remove a key from a dictionary.**