Assignment Code: DA-AG-006

Statistics Advanced - 1| **Assignment**

Instructions: Carefully read each question. Use Google Docs, Microsoft Word, or a similar tool to create a document where you type out each question along with its answer. Save the document as a PDF, and then upload it to the LMS. Please do not zip or archive the files before uploading them. Each question carries 20 marks.

Total Marks: 200

Question 1: What is a random variable in probability theory?

Answer: A random variable is a variable that assigns numerical values to the outcomes of a random experiment.
It helps us study and analyze random phenomena mathematically.
There are two types:

1. Discrete random variable : takes countable values (e.g., rolling a die, number of heads in coin toss).
2. Continuous random variable : takes uncountably infinite values (e.g., height, weight, time).

Question 2: What are the types of random variables?

Answer: There are two types:

1. Discrete random variable : takes countable values (e.g., rolling a die, number of heads in coin toss).
2. Continuous random variable : takes uncountably infinite values (e.g., height, weight, time).

Question 3: Explain the difference between discrete and continuous distributions.

Answer:

Discrete Distribution

- Deals with countable outcomes.
- Probability is assigned to each possible value.
- Example: Number of heads in 3 coin tosses, rolling a die.
- Represented by a Probability Mass Function (PMF).

Continuous Distribution

- Deals with uncountably infinite outcomes (real numbers in an interval).
- Probability of any exact value is 0, but we measure probability over an interval.
- Example: Height of students, time taken to run 100m.
- Represented by a Probability Density Function (PDF).

Question 4: What is a binomial distribution, and how is it used in probability?

Answer: A binomial distribution is a type of probability distribution that models the number of successes in a fixed number of independent trials of a Bernoulli experiment (an experiment with only two outcomes: success or failure).

Question 5: What is the standard normal distribution, and why is it important?

Answer:
A standard normal distribution is a special case of the normal distribution.
It is a bell-shaped, symmetric probability distribution with:

- Mean = 0
- Standard deviation = 1

The random variable that follows it is usually denoted as Z, called the standard normal variable.

Question 6: What is the Central Limit Theorem (CLT), and why is it critical in statistics?

Answer:
The Central Limit Theorem (CLT) states that when you take sufficiently large random samples from a population, regardless of the population's original distribution, the distribution of the sample means will approximate a normal distribution. Central Limit Theorem is crtical because it allows us to use the normal distribution for inference, even when the population distribution is not normal

Question 7: What is the significance of confidence intervals in statistical analysis?

Answer: significance of confidence intervals in statistical analysis

- Provide a range of values for the true population parameter.
- Indicate the level of precision of the estimate.
- Show the reliability of the statistical result.
- Help in making informed decisions under uncertainty.
- Connect directly to hypothesis testing by indicating plausible values.

.

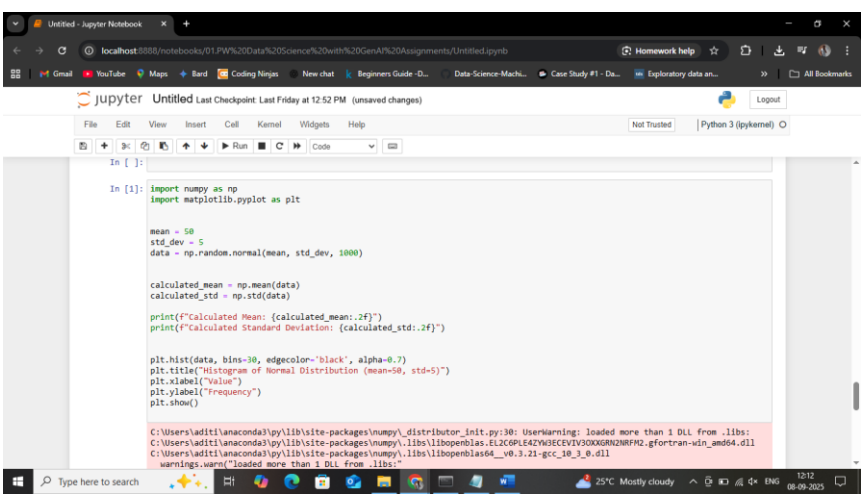
Question 8: What is the concept of expected value in a probability distribution?

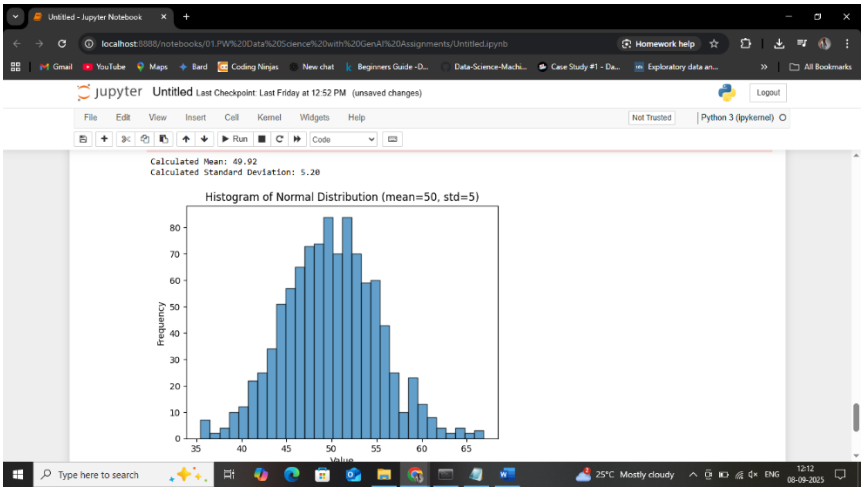
Answer: The expected value of a probability distribution is the long-run average outcome of a random variable when an experiment is repeated many times.
It is also called the mean of the distribution and is calculated as a weighted average of all possible values, where the weights are their respective probabilities.

Question 9: Write a Python program to generate 1000 random numbers from a normal distribution with mean = 50 and standard deviation = 5. Compute its mean and standard deviation using NumPy, and draw a histogram to visualize the distribution.

(*Include your Python code and output in the code box below.*)

Answer:




Untitled - Jupyter Notebook
localhost:8888/notebooks/01.PW%20Data%20Science%20with%20GenAI%20Assignments/Untitled.ipynb
Homework help
jupyter Untitled Last Checkpoint: Last Friday at 12:52 PM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Calculated Mean: 49.92
Calculated Standard Deviation: 5.20
Histogram of Normal Distribution (mean=50, std=5)
Frequency
Value
Type here to search
25°C Mostly cloudy
ENG
12:12
08-09-2025

Question 10: You are working as a data analyst for a retail company. The company has collected daily sales data for 2 years and wants you to identify the overall sales trend.

daily_sales = [220, 245, 210, 265, 230, 250, 260, 275, 240, 255,

235, 260, 245, 250, 225, 270, 265, 255, 250, 260]

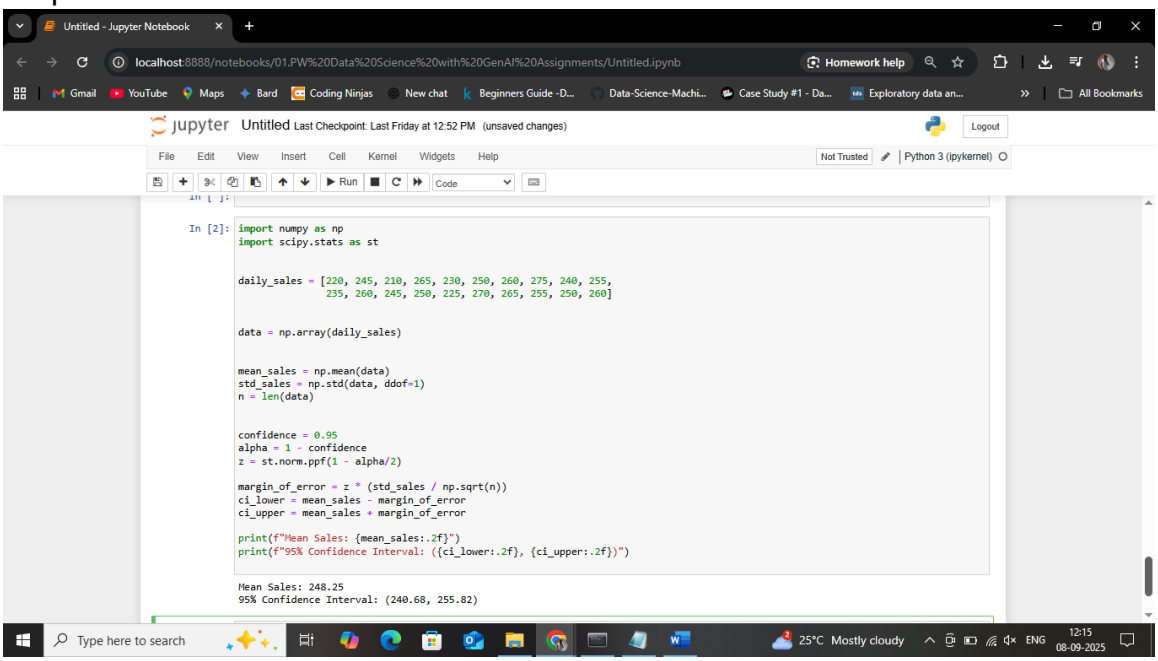
- Explain how you would apply the Central Limit Theorem to estimate the average sales with a 95% confidence interval.
- Write the Python code to compute the mean sales and its confidence

interval. (*Include your Python code and output in the code box below.*)

Answer:
Step 1: Applying the Central Limit Theorem (CLT)

- The population distribution of sales is unknown (may not be normal).
- According to the Central Limit Theorem (CLT):
  The distribution of sample means approaches a normal distribution as sample size increases, regardless of the population's shape.
- This allows us to use the normal distribution to estimate a confidence interval for the mean.

Step 2:



Step 3:
Mean Sales: 248.25
95% Confidence Interval: (240.68, 255.82)

Step 4 :
The **average daily sales** from the sample is **248.25 units**.
With **95% confidence**, the true population mean daily sales lies between **240.68 and 255.82 units**.
This range helps management understand the expected daily sales trend and plan inventory, staffing, and marketing accordingly.