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 numerical outcome of a random process. It maps outcomes of a probabilistic experiment to real numbers.

Question 2: What are the types of random variables?

Answer:

- **Discrete**: Takes countable values (e.g., number of heads in coin tosses).
- **Continuous**: Takes any value within a range (e.g., height, temperature).

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

Answer:

- **Discrete**: Probability is assigned to specific values.
- **Continuous**: Probability is spread over intervals; uses probability density functions.

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

Answer: It models the number of successes in a fixed number of independent Bernoulli trials. Used in scenarios like quality control or survey analysis.

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

Answer: It's a normal distribution with mean 0 and standard deviation 1. It simplifies statistical calculations and is used in z-tests and confidence intervals.

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

Answer: CLT states that the sampling distribution of the sample mean approaches a normal distribution as sample size increases, regardless of the population's distribution. It enables inference using normal models.

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

Answer: Confidence intervals estimate the range within which a population parameter lies, with a certain level of confidence (e.g., 95%). They express uncertainty in estimates.

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

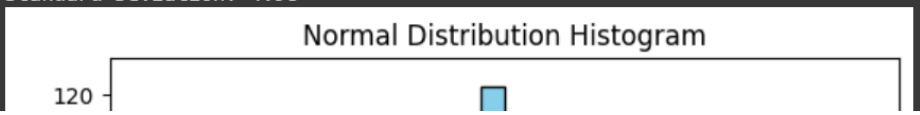
Answer: Expected value is the long-run average of a random variable. It's calculated by summing all possible values weighted by their probabilities.

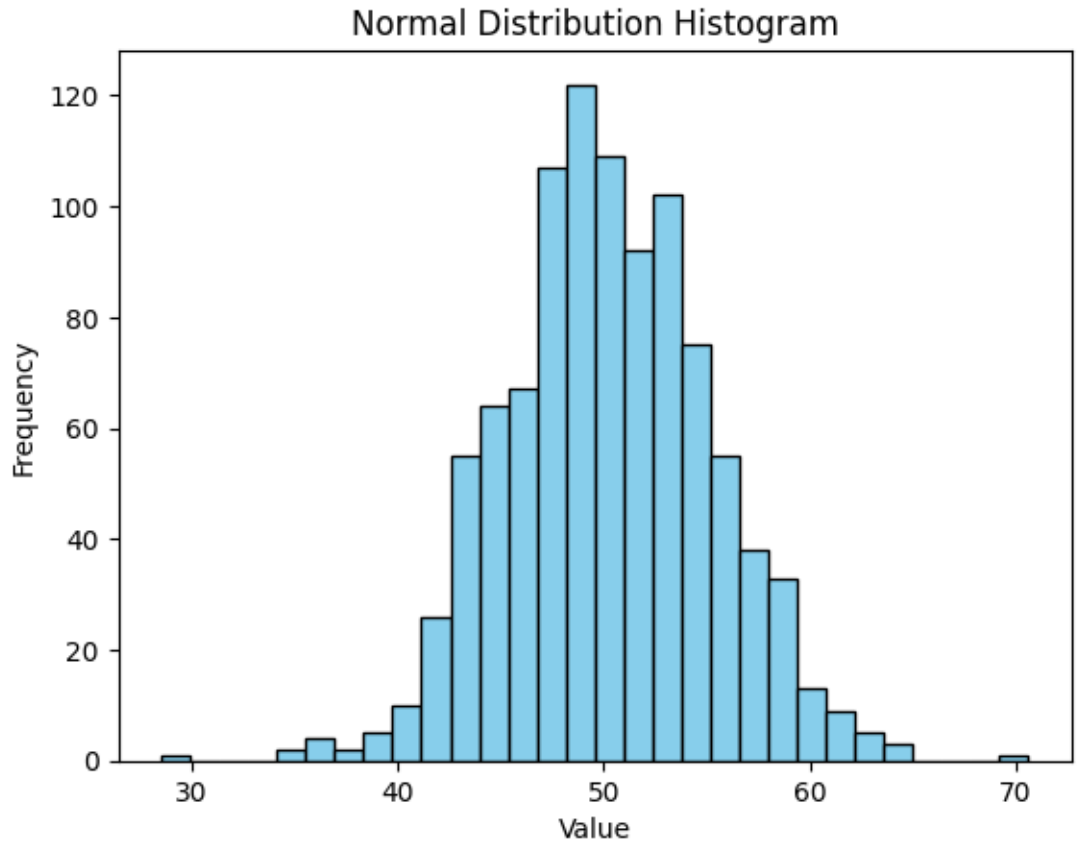
Question 9: Write a Python program to generate 1000 random numbers from a normal distribution with mean = 50 and standard deviation = 5. Compute its mean and standard deviation using NumPy and draw a histogram to visualize the distribution. (Include your Python code and output in the code box below.)

Answer:

```python
import numpy as np
import matplotlib.pyplot as plt
# Generate random numbers
data = np.random.normal(loc=50, scale=5, size=1000)
# Compute mean and std deviation
mean = np.mean(data)
std_dev = np.std(data)
# Output
print(f"Mean: {mean:.2f}")
print(f"Standard Deviation: {std_dev:.2f}")
# Histogram
plt.hist(data, bins=30, color='skyblue', edgecolor='black')
plt.title("Normal Distribution Histogram")
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
```

```
Mean: 50.19
Standard Deviation: 4.98
```





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:

```python
import numpy as np
import scipy.stats as stats
daily_sales = [220, 245, 210, 265, 230, 250, 260, 275, 240, 255,
               235, 260, 245, 250, 225, 270, 265, 255, 250, 260]
# Mean and standard error
mean_sales = np.mean(daily_sales)
std_error = stats.sem(daily_sales)
# 95% confidence interval
ci = stats.norm.interval(0.95, loc=mean_sales, scale=std_error)
print(f"Mean Sales: {mean_sales:.2f}")
print(f"95% Confidence Interval: ({ci[0]:.2f}, {ci[1]:.2f})")
```

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