Question 1(MSQ)

Scenario: A student is writing a Python script to visualize data as a graph and needs to import libraries to handle numerical data and plotting.

Question: Which of the following libraries should the student import? (Select all that apply.)

A) numpy

B) pandas

C) matplotlib.pyplot

D) seaborn

Answer: A, C

Explanation: numpy is essential for numerical computations (e.g., arrays for data points), and matplotlib.pyplot (often as plt) is the core library for creating plots in Python. pandas is useful for data frames but not strictly necessary for basic plotting, and seaborn enhances visuals but isn't required for a simple graph.

Question 2(MSQ)

Scenario: A student creates a Python script to display a graph and wants to ensure it appears with a custom size and is shown on screen.

Approach: Think about common commands to set up and display a plot in Python.

Question: Which of the following commands are needed to set a figure size and display the plot? (Select all that apply.)

A) plt.plot()

B) plt.show()

C) plt.figure(figsize=(8, 5))

D) plt.scatter()

Answer: B, C

Explanation: plt.figure(figsize=(8, 5)) sets the figure size (e.g., 8 inches wide, 5 inches tall), and plt.show() displays the plot. plt.plot() and plt.scatter() draw data (lines or points), but they don't control size or display, making B and C the key setup commands.

Question 3 (Numerical)

Scenario: A researcher calculates the mean number of animals crossing a river daily during a migration week, with counts: 40, 50, 60, 45, 55, 50, 48. They need the average to report trends.

Approach: Add the values and divide by the number of days.

Question: What is the mean number of animals crossing per day? (Round to 1 decimal place.)

Answer: 49.7

Explanation: Mean = (40 + 50 + 60 + 45 + 55 + 50 + 48) / 7 = 348 / 7 ≈ 49.714, rounded to 49.7. This mirrors the mean calculation process, showing how averages summarize data.

Question 4

Scenario: A student wants to save their Python-generated plot as a PNG file to share with classmates after plotting the data.

Approach: Recall the Python command to export a figure to a file.

Question: Which command should they use to save the plot?

A) plt.savefig('plot.png')

B) plt.export('plot.png')

C) plt.save('plot.png')

D) plt.write('plot.png')

Answer: A

Explanation: plt.savefig('plot.png') is the correct matplotlib command to save the current figure as a PNG file. The other options (export, save, write) are not valid matplotlib functions, making A the only correct choice.

Question 5: Identifying Runtime Errors in a Python Plotting Script

Scenario: A student runs a Python script to plot data, but it crashes with a runtime error. They suspect issues with the code feeding the plot.

Approach: Consider common errors that occur **at runtime** when executing a Python plotting script.

Question: Which of the following could cause the runtime error? (Select all that apply.)

- A) Dividing a number by zero in a calculation
- B) Attempting to read a non-existent data file when loading the dataset
- C) Using an undefined variable in the plot command
- D) Misspelling the plot function (e.g., `plt.plt()`)

Correct Answer: A, B, C

Explanation:

- **A) Division by zero** in a calculation (e.g., `x / 0`) raises a **ZeroDivisionError**, stopping execution.
- **B) Reading a non-existent data file** (e.g., `pd.read_csv("missing_file.csv")`) raises a **FileNotFoundError**, which occurs at runtime. Since data visualization often involves loading datasets, this is a relevant runtime issue.
- **C) Using an undefined variable** (e.g., `plt.plot(y)` when `y` is not defined) causes a **NameError**, crashing the script.
- **D) Misspelling** `plt.plt()` would cause a **SyntaxError**, which prevents the script from running at all rather than being a runtime error.

Question 6

Scenario: A student plots a dataset in Python and wants to add a title to make it clearer for their teacher.

Approach: Think about the matplotlib command for adding a title above the plot.

Question: Which command should they use to add a title?

A) plt.xlabel('Title')

B) plt.title('My Plot')

C) plt.legend('Title')

D) plt.text('Title')

Answer: B

Explanation: plt.title('My Plot') adds a title above the plot in matplotlib. plt.xlabel() labels the x-axis, plt.legend() adds a legend (not a title), and plt.text() places text anywhere but isn't specific to titles. B is the correct choice for this purpose.

Question 7

Scenario: A student visualizes animal migration data and notices a histogram with two distinct peaks. They need to classify the distribution type.

Approach: Consider the number of peaks in the histogram.

Question: What type of distribution does the histogram likely represent?

A) Unimodal

B) Bimodal

C) Multimodal

D) Uniform

Answer: B

Explanation: A bimodal distribution is defined as having two peaks (e.g., two migration surges). Unimodal has one peak, multimodal has three or more, and uniform has no peaks (flat). Two peaks indicate B.

Question 8

Scenario: A student writes a Python script to calculate the average rating of a documentary based on the scores: 4, 4, 4, 4, 5, 5, 5, 1, 4. They need to choose the correct library and function.

Approach: Look for a Python library that simplifies numerical calculations.

Question: Which command should they use to compute the mean?

A) import numpy as np; np.mean([4, 4, 4, 4, 5, 5, 5, 1, 4])

B) import scipy.stats as stats; stats.mean([4, 4, 4, 4, 5, 5, 5, 1, 4])

C) import seaborn as sns; sns.mean([4, 4, 4, 4, 5, 5, 5, 1, 4])

D) import matplotlib.pyplot as plt; plt.mean([4, 4, 4, 4, 5, 5, 5, 1, 4])

Answer: A

Explanation: Numpy (imported as np) is used with np.mean() to calculate the mean (e.g., 3.8 for these ratings). SciPy's stats focuses on mode, Seaborn is for visualization, and Matplotlib is for plotting—not mean calculations. A is correct.

Question 9(MSQ)

Scenario: A wildlife expert tracks daily animal crossings during a migration and wants to visualize the data's distribution using a histogram in Python. They need to select the right tools.

Approach: Consider libraries and commands used to create histograms.

Question: Which of the following are required to generate a histogram? (Select all that apply.)

A) import seaborn as sns

B) sns.histplot(data, bins=30)

C) import matplotlib.pyplot as plt

D) plt.show()

Answer: A, B, C, D

Explanation: Histograms are demonstrated with Seaborn (sns.histplot(data, bins=30)) and Matplotlib (plt.show() to display). Both import seaborn as sns and import matplotlib.pyplot as plt are needed, as Seaborn builds on Matplotlib. All four are essential for the full process.

Question 10(MSQ)

Scenario: A data analyst compares test scores of two classes using box plots and notices differences in spread and skewness. They want to interpret the plots' components.

Approach: Recall the parts of a box plot and what they indicate.

Question: Which of the following can a box plot reveal about the scores? (Select all that apply.)

A) The median score of each class

B) The range of the middle 50% of scores

C) The most frequent score (mode)

D) Outliers beyond the typical range

Answer: A, B, D

Explanation: A box plot shows the median (A, line in the box), interquartile range (B, IQR, middle 50%), and outliers (D, points beyond whiskers). It doesn't display the mode (C), which requires a histogram or frequency count, making C incorrect.

Note: Q11 to Q15 are very specific to Lecture on 11th March and Notes on Statistical Hypothesis Testing. Lecture Link

Question 11(MSQ)

Scenario: A researcher is studying why a new café failed to attract customers despite an average group interest threshold of 1.8. They need to identify key statistical insights.

Approach: Think about the Silent Café Paradox and its lessons on traditional statistics.

Question: Which of the following are limitations of using only mean, median, and mode in this case? (Select all that apply.)

A) They ignore activation sequences

B) They miss the need for a trendsetter (0-threshold)

C) They don't account for sample size

D) They overlook threshold dependencies

Answer: A, B, D

Explanation: The Silent Café Paradox shows that mean (1.8) misleads because it ignores activation chains (A), the need for a 0-threshold initiator (B), and threshold dependencies (D). Sample size (C) isn't highlighted as a primary issue here, so it's not selected.

Question 12

Scenario: A student wants to test if two companies' salaries differ significantly and needs to import the right Python tool for a T-test.

Approach: Look for the library used for statistical tests like T-tests.

Question: Which import statement should they use?

A) import numpy as np

B) import scipy.stats as stats

C) import seaborn as sns

D) import matplotlib.pyplot as plt

Answer: B

Explanation: Use scipy.stats (aliased as stats) for tests like ttest_ind (T-test). NumPy handles arrays, Seaborn is for visualization, and Matplotlib is for plotting—not hypothesis testing. B is correct.

Question 13

Scenario: A student interprets a p-value of 0.03 from a T-test comparing two groups' performance and needs to decide on the null hypothesis.

Approach: Compare the p-value to the significance level (α).

Question: What should they conclude if the significance level is 0.05?

A) Fail to reject the null hypothesis

B) Reject the null hypothesis

C) Increase the significance level

D) Recalculate the p-value

Answer: B

Explanation: The lecture states: if $p<0.05$, reject H0 H_0 H0 (no difference). Here, $0.03<0.05$ $0.03 < 0.05$ $0.03<0.05$, indicating a significant difference. A is wrong (p is below α), C and D are unnecessary steps.

Question 14(MSQ)

Scenario: A sociologist studies group behavior in four friend groups and runs hypothesis tests to compare activation rates. They need to avoid errors in interpretation.

Approach: Consider the types of errors in hypothesis testing from the PDF's cheat sheet.

Question: Which of the following are potential errors they should watch for? (Select all that apply.)

A) Type I Error (false positive)

B) Type II Error (false negative)

C) Using a Z-test for a small sample

D) Ignoring variance checks before a T-test

Answer: A, B, C, D

Explanation: (A) Type I (false positive, 5% risk), (B) Type II (false negative), (C) Z-test misuse (should be n>30 n > 30 n>30), and (D) skipping Levene's Test before T-tests. All are critical pitfalls in hypothesis testing.

Question 15(NAT)

Scenario: A manager tests if phone preference (Apple: 30M/40F, Samsung: 50M/30F, OnePlus: 20M/30F) is independent of gender using a Chi-square test. They calculate the degrees of freedom (dof).

Approach: Use

$$dof = (rows - 1) \times (columns - 1)$$

for a contingency table.

Question: What is the degrees of freedom for this Chi-square test?

Answer: 2

Explanation: For a 3x2 table (3 phone types, 2 genders),

$$dof = (3 - 1) \times (2 - 1) = 2 \times 1 = 2$$

This is how we test independence and table structure understanding using Chi-square.