1. Explain the properties of the F-distribution.

ANSWER - **Properties of the F-Distribution:**

1. **Asymmetrical**: The F-distribution is not symmetric; it is skewed to the right.
2. **Starts at Zero**: Values are always ≥0it never goes negative.
3. **Shape Depends on Degrees of Freedom**:
    - Two parameters: df1 (numerator) and df2df_2df2 (denominator).
    - Higher df1 and df2 make the distribution more symmetric.
4. **Used for Variance Ratios**:
    - Commonly used in ANOVA and hypothesis testing.
5. **Area Under the Curve**:
    - Total probability under the curve is 1.

Q2). In which types of statistical tests is the F-distribution used, and why is it appropriate for these tests?

ANSWER - **Statistical Tests Using F-Distribution:**

1. **ANOVA (Analysis of Variance)**:
    - Compares the variances between groups to test if their means differ.
    - **Why**: F-distribution handles variance ratios.
2. **Regression Analysis**:
    - Tests the overall significance of the regression model.
    - **Why**: Evaluates variance explained by predictors vs. error variance.
3. **Equality of Variances (F-Test)**:
    - Compares variances of two datasets.
    - **Why**: F-distribution works for variance ratios.

Why Appropriate:

The F-distribution models ratios of variances and handles right-skewed data well.

Q3) 3. What are the key assumptions required for conducting an F-test to compare the variances of two populations?

ANSWER - **Key Assumptions for F-Test:**

1. **Independence**: Samples from both populations must be independent.
2. **Normality**: Data in both populations should follow a normal distribution.

3. **Random Sampling**: Samples must be randomly selected from the populations.
4. **Positive Variances**: Variances of the populations must be greater than zero.

These assumptions ensure the F-test is valid and results are reliable.

Q4) 4. What is the purpose of ANOVA, and how does it differ from a t-test?

ANSWER - **Purpose of ANOVA:**

- ANOVA (Analysis of Variance) tests if the **means of three or more groups** are significantly different.
- It analyzes **variances** within groups and between groups to determine differences.

Difference from a t-Test:

1. **Number of Groups**:
    - **t-Test**: Compares the means of **two groups**.
    - **ANOVA**: Compares the means of **three or more groups**.
2. **Risk of Error**:
    - Using multiple t-tests increases the chance of **Type I error** (false positives).
    - ANOVA handles multiple comparisons in one test, reducing this risk.

ANOVA is ideal when comparing more than two groups!

Q5) 5. Explain when and why you would use a one-way ANOVA instead of multiple t-tests when comparing more than two groups.

ANSWER - **When to Use One-Way ANOVA:**

- Use one-way ANOVA when comparing **means of three or more groups** (e.g., test scores of students in 3 different classes).

Why Use ANOVA Instead of Multiple t-Tests:

1. **Controls Error**: Multiple t-tests increase the risk of **Type I error** (false positives). ANOVA controls this by performing one overall test.
2. **Efficiency**: ANOVA compares all groups simultaneously, saving time and effort.
3. **Appropriate for Variance**: ANOVA evaluates differences using group variances, making it statistically robust.

In short, ANOVA is **more accurate and reliable** for comparing multiple groups.

Q6) 6. Explain how variance is partitioned in ANOVA into between-group variance and within-group variance. How does this partitioning contribute to the calculation of the F-statistic?

ANSWER - **Variance Partitioning in ANOVA:**

1. **Total Variance**: The overall variability in the data.
    - It is split into:
        - **Between-Group Variance**: Variability due to differences between group means.
        - **Within-Group Variance**: Variability within each group (individual differences).

$$F = \frac{\text{Between-Group Variance (MSB)}}{\text{Within-Group Variance (MSW)}}$$

    - **MSB**: Mean Square Between groups (variance explained by groups).
    - **MSW**: Mean Square Within groups (unexplained variance).
2. **High F-Value**: Indicates group means are significantly different relative to within-group variation.

Partitioning variance allows ANOVA to test if group differences are real or due to random variation.

Q7) 7. Compare the classical (frequentist) approach to ANOVA with the Bayesian approach. What are the key differences in terms of how they handle uncertainty, parameter estimation, and hypothesis testing?

ANSWER - The main differences between frequentist and Bayesian ANOVA are:

1. **Uncertainty**:
    - **Frequentist**: Uses p-values to show how likely the data is under the null hypothesis.
    - **Bayesian**: Uses probability distributions to show uncertainty about the parameters.
2. **Parameter Estimation**:
    - **Frequentist**: Estimates parameters as fixed values (e.g., sample means).
    - **Bayesian**: Estimates parameters as random variables with a distribution (posterior).
3. **Hypothesis Testing**:
    - **Frequentist**: Tests hypotheses with p-values.

- **Bayesian**: Compares models using Bayes factors, showing which hypothesis is more likely.

Q8) Question: You have two sets of data representing the incomes of two different professions Profession A: [48, 52, 55, 60, 62 Profession B: [45, 50, 55, 52, 47] Perform an F-test to determine if the variances of the two professions' incomes are equal. What are your conclusions based on the F-test? Task: Use Python to calculate the F-statistic and p-value for the given data. Objective: Gain experience in performing F-tests and interpreting the results in terms of variance comparison.

ANSWER - https://github.com/Abhishek-D8mik3/Assignments

Q9) 9. Question: Conduct a one-way ANOVA to test whether there are any statistically significant differences in average heights between three different regions with the following data Region A: [160, 162, 165, 158, 164 Region B: [172, 175, 170, 168, 174 Region C: [180, 182, 179, 185, 183 Task: Write Python code to perform the one-way ANOVA and interpret the results Objective: Learn how to perform one-way ANOVA using Python and interpret F-statistic and p-value.

ANSWER - https://github.com/Abhishek-D8mik3/Assignments