1. **Explain the properties of the F-distribution.**

Ans :

The F distribution is a probability distribution that is useful in the context of comparing Variance of two or more sample

It is Right Skewed and takes only non negative values

F stats = S1^2/S2^2

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

The F-distribution is commonly used in statistical tests that compare variances between groups, specifically in Analysis of Variance (ANOVA), regression analysis, and tests of equal variances (like the Levene's test).

ANOVA (Analysis of Variance):

In ANOVA, the F-test is used to compare the variances between group means to assess if there are statistically significant differences among them. The F-distribution is suitable here because ANOVA involves comparing the variability of group means against the variability within the groups.

It measures how much the group means deviate from the overall mean relative to the within-group variability.

Regression Analysis:

In regression, the F-test checks the overall significance of the model by comparing the explained variance (variance in the outcome variable explained by the predictors) to the unexplained variance (residual error).

The F-distribution is ideal because it provides a framework for testing the null hypothesis that none of the independent variables have a statistically significant relationship with the dependent variable.

Tests of Equal Variances (e.g., Levene's Test):

F-tests can also be used to determine if the variances between two or more groups are equal, which is an assumption in many parametric tests.

Since the F-distribution is derived from the ratio of two independent chi-square distributions, it is appropriate for comparing sample variances.

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

The Population from which samples are drawn should be normally distributed

The sample should be independent from each other / Random

Absence of outlier

Homogenety of variance

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

The ANOVA is a Stats method used to compare the means of 2 or more groups

Using Variance , you want to determine , if the mean value of two or more than two groups are same or different

Generally ANOVA is used for more than 2 groups

Because for two groups we can use t test

T test basically used when the sample size is less than 30 and or there is no population standard deviation

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

Conducting multiple t-tests on each pair of group means increases the risk of a Type I error (i.e., incorrectly rejecting the null hypothesis when it is true). For example, if you have three groups (A, B, and C), comparing each pair (A vs. B, A vs. C, B vs. C) requires three t-tests, each with its own 5% risk of error. This cumulative error risk grows with each additional test.

A one-way ANOVA performs a single test for overall differences among the group means, thus maintaining the desired error rate (usually 5%) across all comparisons.

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?

For the Between group variance and within group variance of ANOVA .

ANOVA **btw the group variance**

Sum of square btw the group

Then the Degree of freedom btw the groups = No: of groups – 1

Mean square btw the group (MSB) = Sum of square btw the group / dof

ANOVA **within the groups**

Sum of square within the group

Then the Degree of freedom within the group = Total no of sample – No of groups

Mean square within the group (MSW) = sum of square within / dof

Fstats = MSB / MSW

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?

Classical (Frequentist) Approach:

In the frequentist framework, uncertainty is addressed by assuming fixed but unknown parameters. The observed data is used to estimate these parameters, with confidence intervals indicating the range in which the parameter is likely to fall, based on repeated sampling.

The approach does not treat parameters as random; instead, it assumes the data varies based on the sampling process.

Bayesian Approach:

In the Bayesian framework, parameters are considered random variables with prior distributions representing our uncertainty about them before observing the data.

After observing data, the Bayesian approach updates the prior distribution with the likelihood of the observed data to produce a posterior distribution, which fully quantifies the uncertainty about parameters.

8. Question: You have two sets of data representing the incomes of two different professions1 V Profession A: [48, 52, 55, 60, 62' V 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 on jupyter notebook

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 data1 V Region A: [160, 162, 165, 158, 164' V Region B: [172, 175, 170, 168, 174' V Region C: [180, 182, 179, 185, 183' V Task: Write Python code to perform the one-way ANOVA and interpret the results

V Objective: Learn how to perform one-way ANOVA using Python and interpret F-statistic and p-value.

Answer on Jupyter notebook