

**Q1. Write a Python function that takes in two arrays of data and calculates the F-value for a variance ratio test. The function should return the F-value and the corresponding p-value for the test.**

In [2]:

```
import numpy as np
from scipy.stats import f

def variance_ratio_test(x, y):
    var_x = np.var(x)
    var_y = np.var(y)
    F = var_x / var_y

    df_x = len(x) - 1
    df_y = len(y) - 1

    p_value = f.cdf(F, df_x, df_y)

    return F, p_value
x = [1, 2, 3, 4, 5]
y = [2, 4, 6, 8, 10]

F, p_value = variance_ratio_test(x, y)

print("F-value:", F)
print("p-value:", p_value)
```

F-value: 0.25  
p-value: 0.10400000000000002

**Q2. Given a significance level of 0.05 and the degrees of freedom for the numerator and denominator of an F-distribution, write a Python function that returns the critical F-value for a two-tailed test.**

In [3]:

```
from scipy.stats import f

def critical_f_value(dfn, dfd, alpha=0.05):
    f_value = f.ppf(1 - alpha / 2, dfn, dfd)
    return f_value

dfn = 3
dfd = 20
alpha = 0.05
critical_f = critical_f_value(dfn, dfd, alpha)
print("Critical F-value:", critical_f)
```

Critical F-value: 3.8586986662732143

**Q3. Write a Python program that generates random samples from two normal distributions with known variances and uses an F-test to determine if the variances are equal. The program should output the F- value, degrees of freedom, and p-value for the test.**

In [6]:

```

import numpy as np
from scipy.stats import f

np.random.seed(123)

n1 = 50
n2 = 75
mean1 = 10
mean2 = 12
var1 = 4
var2 = 6
sample1 = np.random.normal(loc=mean1, scale=np.sqrt(var1), size=n1)
sample2 = np.random.normal(loc=mean2, scale=np.sqrt(var2), size=n2)

F = np.var(sample1, ddof=1) / np.var(sample2, ddof=1)
dfn = n1 - 1
dfd = n2 - 1
p_value = f.sf(F, dfn, dfd)

print("F-value:", F)
print("Degrees of freedom numerator:", dfn)
print("Degrees of freedom denominator:", dfd)
print("p-value:", p_value)

```

F-value: 0.7792593732703956  
 Degrees of freedom numerator: 49  
 Degrees of freedom denominator: 74  
 p-value: 0.8229821876353054

**Q4. The variances of two populations are known to be 10 and 15. A sample of 12 observations is taken from each population. Conduct an F-test at the 5% significance level to determine if the variances are significantly different.**

Define the null hypothesis and alternative hypothesis: Null hypothesis ( $H_0$ ): The variances of the two populations are equal. Alternative hypothesis ( $H_A$ ): The variances of the two populations are significantly different. Determine the significance level ( $\alpha$ ) and degrees of freedom for the numerator and denominator of the F-distribution: Significance level ( $\alpha$ ) = 0.05 Degrees of freedom numerator (dfn) = sample size of population 1 - 1 = 12 - 1 = 11 Degrees of freedom denominator (dfd) = sample size of population 2 - 1 = 12 - 1 = 11 Calculate the F-statistic using the formula:  $F = s_1^2 / s_2^2$

where  $s_1^2$  and  $s_2^2$  are the sample variances for population 1 and population 2, respectively. Using the given information, we can conduct an F-test as follows:  $H_0: \sigma_1^2 = \sigma_2^2$   $H_A: \sigma_1^2 \neq \sigma_2^2$

$\alpha = 0.05$ , dfn = 11, dfd = 11

Calculate the F-statistic:  $F = s_1^2 / s_2^2$

Since the variances are known to be 10 and 15 for population 1 and population 2, respectively, we can use these values as the sample variances:  $F = 10 / 15 = 0.67$

Find the p-value for the F-statistic using the cumulative distribution function (cdf) of the F-distribution with degrees of freedom dfn and dfd:  $p\text{-value} = 2 * \min(f.cdf(F, dfn, dfd), 1 - f.cdf(F, dfn, dfd))$

Here we multiply the cdf by 2 because it is a two-tailed test.  $p\text{-value} = 2 * \min(f.cdf(0.67, 11, 11), 1 - f.cdf(0.67, 11, 11)) = 2 * \min(0.255, 0.282) = 0.51$

Compare the p-value with the significance level  $\alpha$ . Since the p-value (0.51) is greater than  $\alpha$  (0.05), fail to reject the null hypothesis and conclude that there is not enough evidence to suggest that the variances of the two populations are different. Therefore, we can conclude that at the 5% significance level, there is no significant difference in the variances of the two populations.

**Q5. A manufacturer claims that the variance of the diameter of a certain product is 0.005. A sample of 25 products is taken, and the sample variance is found to be 0.006. Conduct an F-test at the 1% significance level to determine if the claim is justified.**

The hypothesis test we will conduct is: Null hypothesis: The variance of the diameter of the product is 0.005. Alternative hypothesis: The variance of the diameter of the product is greater than 0.005. We will use the F-test to compare the sample variance with the claimed population variance:

$F = \text{sample variance} / \text{population variance}$

Under the null hypothesis, the F statistic follows an F distribution with degrees of freedom (n-1) for the numerator and (m-1) for the denominator, where n is the sample size and m is the assumed population size. In this case, we have: n = 25 (sample size) m = infinity (population size is not specified) sample variance = 0.006 claimed population variance = 0.005 Therefore, the F statistic is:  $F = 0.006 / 0.005 = 1.2$

The critical value of F at the 1% level of significance with (24, infinity) degrees of freedom is 2.75. Since the calculated F value of 1.2 is less than the critical value of 2.75, we fail to reject the null hypothesis.

Therefore, we do not have sufficient evidence to conclude that the population variance is greater than 0.005. The manufacturer's claim is justified at the 1% level of significance.

**Q6. Write a Python function that takes in the degrees of freedom for the numerator and denominator of an F-distribution and calculates the mean and variance of the distribution. The function should return the mean and variance as a tuple.**

In [7]:

```
import scipy.stats as stats

def f_distribution_mean_variance(df_numerator, df_denominator):

    mean = df_denominator / (df_denominator - 2)
    variance = (2 * (df_denominator**2) * (df_numerator + df_denominator - 2)) / ((df_nu
    return (mean, variance)

f_distribution_mean_variance(5, 10)
```

Out[7]:

```
(1.25, 1.3541666666666667)
```

**Q7. A random sample of 10 measurements is taken from a normal population with unknown variance. The sample variance is found to be 25. Another random sample of 15 measurements is taken from another normal population with unknown variance, and the sample variance is found to be 20. Conduct an F-test at the 10% significance level to determine if the variances are significantly different.**

In [8]:

```

import scipy.stats as stats

# Define the sample variances and sample sizes
s1 = 25
s2 = 20
n1 = 10
n2 = 15

# Calculate the F statistic
F = s1 / s2

# Calculate the p-value
p_value = stats.f.cdf(F, n1-1, n2-1)

# Calculate the critical values of F
alpha = 0.1
f_critical_lower = stats.f.ppf(alpha/2, n1-1, n2-1)
f_critical_upper = stats.f.ppf(1-alpha/2, n1-1, n2-1)

# Print the results
print("F statistic:", F)
print("p-value:", p_value)
print("Critical values of F:", f_critical_lower, f_critical_upper)

if F < f_critical_lower or F > f_critical_upper:
    print("Reject the null hypothesis")
else:
    print("Fail to reject the null hypothesis")

```

```

F statistic: 1.25
p-value: 0.6583902808707023
Critical values of F: 0.3305268601412525 2.6457907352338195
Fail to reject the null hypothesis

```

**Q8. The following data represent the waiting times in minutes at two different restaurants on a Saturday night: Restaurant A: 24, 25, 28, 23, 22, 20, 27; Restaurant B: 31, 33, 35, 30, 32, 36. Conduct an F-test at the 5% significance level to determine if the variances are significantly different.**

We will conduct an F-test to determine if the variances of the waiting times at the two restaurants are significantly different. The null hypothesis is that the variances are equal, and the alternative hypothesis is that the variances are not equal. First, we calculate the sample variances for each restaurant: Restaurant A:  $s_1^2 = ((24-25.71)^2 + (25-25.71)^2 + \dots + (27-25.71)^2) / 6 = 7.61$  Restaurant B:  $s_2^2 = ((31-32.83)^2 + (33-32.83)^2 + \dots + (36-32.83)^2) / 5 = 5.5$  Next, we calculate the F-statistic:  $F = s_1^2 / s_2^2 = 7.61 / 5.5 = 1.38$

The degrees of freedom for the numerator and denominator are  $6-1 = 5$  and  $5-1 = 4$ , respectively. Using a significance level of 0.05, the critical values for an F-distribution with (5,4) degrees of freedom are 0.063 and 5.987. Since the calculated F value of 1.38 is between the critical values of 0.063 and 5.987, we fail to reject the null hypothesis. Therefore, we do not have sufficient evidence to conclude that the variances of the waiting times at the two restaurants are significantly different at the 5% level of significance.

**Q9. The following data represent the test scores of two groups of students: Group A: 80, 85, 90, 92, 87, 83; Group B: 75, 78, 82, 79, 81, 84. Conduct an F-test at the 1% significance level to determine if the variances are significantly different. import numpy as np**

In [10]:

```
import numpy as np
import scipy.stats as stats

A = np.array([80, 85, 90, 92, 87, 83])
B = np.array([75, 78, 82, 79, 81, 84])

s1_squared = np.var(A, ddof=1)
s2_squared = np.var(B, ddof=1)

F = s1_squared / s2_squared

p_value = stats.f.cdf(F, len(A)-1, len(B)-1)

alpha = 0.01
f_critical_lower = stats.f.ppf(alpha/2, len(A)-1, len(B)-1)
f_critical_upper = stats.f.ppf(1-alpha/2, len(A)-1, len(B)-1)

print("F statistic:", F)
print("p-value:", p_value)
print("Critical values of F:", f_critical_lower, f_critical_upper)

if F < f_critical_lower or F > f_critical_upper:
    print("Reject the null hypothesis")
else:
    print("Fail to reject the null hypothesis")
```

F statistic: 1.9442622950819677

p-value: 0.7584478225464656

Critical values of F: 0.066936171954696 14.939605459912224

Fail to reject the null hypothesis