```python
import pandas as pd
import numpy as np
```

The Fishers (F) Distribution

Much like the normal and chi-squared distributions have unique applications, the F distribution has its own role. It's particularly relevant when you're comparing variances or testing hypotheses related to variances. In this distribution, you compute an f statistic, similar to how you'd calculate a z-score in the normal distribution or a chi-squared statistic in the chi-squared distribution. The F statistic is then used to find probabilities associated with specific values, just like the z-score in the normal distribution or the chi-squared statistic in the chi-squared distribution. This notebook is a demonstrates the process of calculating an f statistic and determining the probabilities linked to particular f statistic values.

The F statistic, sometimes referred to as an F value, is a stochastic variable following an F distribution. (We will delve deeper into the F distribution in the following section.)

To calculate an F statistic, you can follow these steps:

1. Draw a random sample of size n1 from a normally distributed population with a standard deviation σ1.
2. Draw a separate, independent random sample of size n2 from another normally distributed population with a standard deviation σ2.
3. The F statistic is computed as the ratio of

> F = s1^2/σ1^2 and s2^2/σ2^2

Where;

F: This represents the F statistic, which is a test statistic used to compare the variances of two or more groups or populations. It's particularly useful for assessing whether the variances are equal (homoscedasticity) or not (heteroscedasticity).

s1^2: This is the sample variance of the first group. In other words, it's the variance of the data from the first sample.

σ1^2: This is the population variance of the first group. It represents the theoretical variance of the entire population from which the first sample is drawn. In practice, this is often unknown, and you estimate it based on the sample data.

s2^2: Similarly, this is the sample variance of the second group.

σ2^2: This is the population variance of the second group. As with σ1^2, it's the theoretical variance of the entire population from which the second sample is drawn and is typically estimated from the sample data

The following sets of equivalent equations are frequently employed to calculate an F statistic:

- $f = [s12/\sigma12] / [s22/\sigma22]$
- $f = [s12 * \sigma22] / [s22 * \sigma12]$
- $f = [(X^2)1 / v1] / [(X^2)2 / v2]$
- $f = [(X^2)1 * v2] / [(X^2)2 * v1]$"

These equations offer different mathematical expressions for calculating the F statistic, but they are equivalent and serve the same purpose in statistical analysis.

Where:

$\sigma1$ = the standard deviation of population 1.

s1 = the standard deviation of the sample taken from population 1.

$\sigma2$ = the standard deviation of population 2.

s2 = the standard deviation of the sample collected from population 2.

(X^2)1 represents the chi-square statistic for the sample obtained from population 1.

v1 indicates the degrees of freedom associated with (X^2)1, and it's calculated as v1 = n1 - 1.

(X^2)2 signifies the chi-square statistic for the sample acquired from population 2.

v2 represents the degrees of freedom associated with (X^2)2, and it's calculated as v2 = n2 - 1.

Note that the degrees of freedom for (X^2)1 and (X^2)2 are determined by subtracting 1 from the respective sample sizes: v1 = n1 - 1 and v2 = n2 - 1

The F Distribution

The F distribution is the probability distribution encompassing all potential values of the f statistic. It has associated degrees of freedom denoted as $v1 = n1 - 1$ and $v2 = n2 - 1$.

The shape of the F distribution is contingent on the degrees of freedom, v1 and v2. In the context of describing an F distribution, the degrees of freedom linked to the standard deviation in the numerator of the f statistic are consistently stated first. Therefore, f(5, 9) indicates an F distribution with v1 = 5 and v2 = 9 degrees of freedom, while f(9, 5) signifies an F distribution with v1 = 9 and v2 = 5 degrees of freedom. It's important to note that the curve represented by f(5, 9) differs from the curve represented by f(9, 5).

The F distribution exhibits several key properties:

The mean of the distribution is v2 / (v2 - 2) for v2 > 2. The variance is [2 * v2^2 * (v1 + v1 - 2)] / [v1 * (v2 - 2)^2 * (v2 - 4)] for v2 > 4.

Cumulative Probability and the F Distribution

Each f statistic corresponds to a unique cumulative probability, representing the likelihood that the f statistic is less than or equal to a specific value.

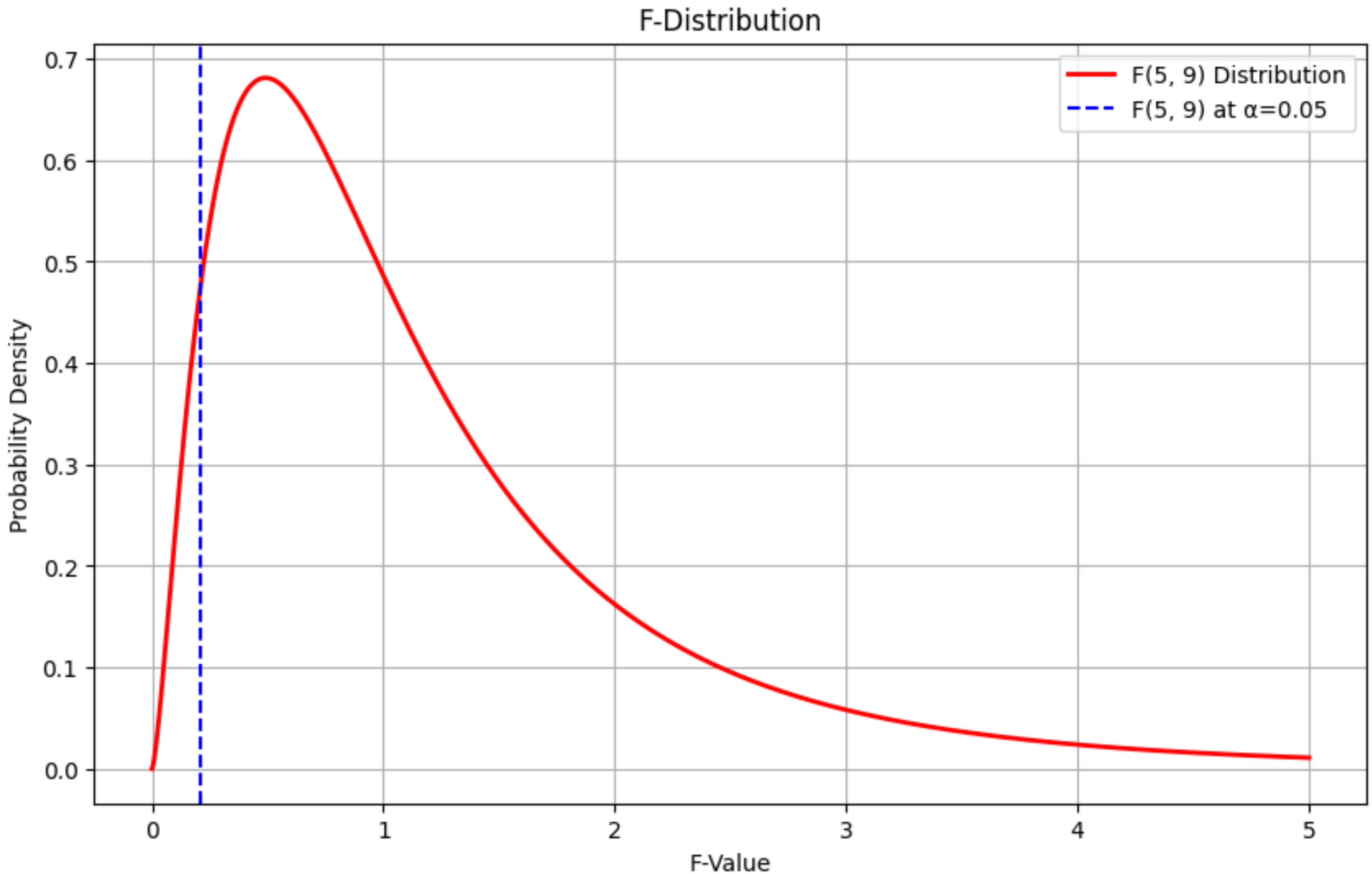
Statisticians use fα to denote the value of an f statistic with a cumulative probability of (1 - α). For example, if we were interested in an f statistic with a cumulative probability of 0.95, we would refer to it as f0.05, as (1 - 0.95) equals 0.05.

To determine the value of fα, you must know the degrees of freedom, v1 and v2. Notationally, degrees of freedom are indicated in parentheses as follows: fα(v1,v2). For instance, f0.05(5, 7) refers to the value of the f statistic with a cumulative probability of 0.95, v1 = 5 degrees of freedom, and v2 = 7 degrees of freedom.

To find the probability linked to a specific f statistic, many statistical resources include tables of F distribution probabilities. Additionally, many graphing calculators have the capability to compute f statistics. On this website, we utilize Stat Trek's F Distribution Calculator, as demonstrated below in Problem 2

```
# Line 1: Import necessary libraries
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

# Line 2: Define degrees of freedom (v1 and v2)
v1 = 5
v2 = 9

# Line 3: Generate a range of F-values
x = np.linspace(0, 5, 1000)

# Line 4: Calculate the probability density functi
pdf = stats.f.pdf(x, v1, v2)

# Line 5: Create a figure for the plot
plt.figure(figsize=(10, 6))

# Line 6: Plot the F-distribution
plt.plot(x, pdf, 'r-', lw=2, label=f'F({v1}, {v2})
plt.title('F-Distribution')
plt.xlabel('F-Value')
plt.ylabel('Probability Density')
plt.legend()
plt.grid(True)

# Line 7: Find and plot a specific probability (cu
```

```
alpha = 0.05  # For example, let's find the F-valu
f_alpha = stats.f.ppf(alpha, v1, v2)
plt.axvline(f_alpha, color='b', linestyle='--', la

plt.legend()

# Line 8: Show the plot
plt.show()
```


F-Distribution
F(5, 9) Distribution
F(5, 9) at α=0.05
Probability Density
F-Value
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
2
3
4
5

The F-distribution has various practical applications:

1. Analysis of Variance (ANOVA): It's used to compare means across multiple groups. For example, in medical research, you might compare the effectiveness of different treatments on a particular condition.

Regression Analysis: In multiple regression analysis, F-tests are used to evaluate the overall significance of the regression model.

Quality Control: F-tests are applied in quality control to determine whether the variances in production processes are consistent.

Finance: In finance, F-tests can be used to compare the volatility of different investment portfolios.

Summary

Fisher's distribution, also known as the F-distribution, is a statistical probability distribution that's commonly used in the analysis of variance and regression. In a mixed-audience-friendly way, think of it as a tool to compare the variances of two sets of data.

For example, if you're analyzing the performance of two different groups (say, Group A and Group B) and you want to know if the variation within these groups is significantly different, you can use the F-distribution. It helps determine if the differences you observe are statistically significant or if they could have occurred by random chance.

In finance, understanding this distribution can be helpful when assessing portfolio risk. For instance, you might want to know if the risk in one investment portfolio is significantly different from another.

In biology, this distribution can be used in experimental design to analyze the variations between groups, like drug test results on two sets of subjects.

So, in a nutshell, Fisher's distribution is like a magnifying glass for spotting differences and assessing their significance in various fields, making it a valuable tool for data analytics and scientific research.

Additional Reading Resources

wolfram

Uregina

NIST

Openstax

Open Library