Statistical Inference

Test for equality of variances

F-test for equality of variances

- F test is used to test the equality of two population variances.
- Testing equality of variances is the prerequisite for many statistical test (like Independent sample t-test).
- Under H0 $\sigma_1^2 = \sigma_2^2$

  Where σ_1^2 and σ_2^2 are the first and second population variances, respectively.

Assumptions for F-test

- The assumptions of F-test are listed below:
  - Random sampling from a defined population
    (employees are selected at random from the company)

  - Population of the testing variable is normally distributed
    (Time taken to complete MIS report should be normally distributed).

- Note that, generally F test is used to validate assumption of equal variance while performing t test for equality of means. The parent population is assumed to follow normal distribution.

Case Study

To execute Test for Equality of Variance in Python, we shall consider the below case as an example.

Background

The company is analysing time to complete MIS report between two groups of employees.
Group I: Experience (0-1 years)
Group II: Experience(1-2 years)

Objective

To test the equality of the variances in time taken to complete MIS in two groups of employees.

Sample Size

Sample size: 14
Variables: time_g1, time_g2

Data Snapshot

F test for 2 variances

Variables

time_g1	time_g2
85	83
95	85
105	96

servations

Columns	Description	Type	Measurement	Possible values
time_g1	Time to complete MIS report by group1	Numeric	Hours	Positive Values
time_g2	Time to complete MIS report by group2	Numeric	Hours	Positive Values

F-test

Testing equality of variances in two samples.

Objective	To test the **equality** of the variances in time taken to complete MIS in two groups of employees.

Null Hypothesis (H_0): Variances of time are equal in two groups. i.e. $\sigma_1^2 = \sigma_2^2$.
Alternate Hypothesis (H_1): Alternative Hypothesis H1: $\sigma_1^2 \neq \sigma_2^2$

Test Statistic	$F = \frac{s_1^2}{s_2^2} \sim F_{\alpha, n_1-1, n_2-1}$ **Where s_1^2 is the sample variance of first sample and, s_2^2 is the sample variance of the second sample. n_1 and n_2 are sample sizes of first and second sample respectively.** **the** **of**
Decision Criteria	Reject the null hypothesis if p-value < 0.05

Computation

	Group I	Group II
Sample Size	$n_1=12$	$n_2=14$
Mean	$\bar{x}_1 = 93.5833$	
Sample Variance	$s_1^2 = 41.9015$	$s_2^2 = 27.1484$
F Value	$F = \frac{s_1^2}{s_2^2}$	1.5434

F-test in Python

```
# Import data
data = pd.read_csv('F test for 2 variances.csv')

# Variance test
import numpy as np
from scipy import stats

x = np.array(data.dropna()['time_g1'])
y = np.array(data['time_g2'])

f = np.var(x, ddof=1)/np.var(y, ddof=1) #calculate F test
statistic
dfn = x.size-1 #define degrees of freedom numerator
dfd = y.size-1 #define degrees of freedom denominator
p = 2*(1-stats.f.cdf(f, dfn, dfd)) #find p-value of F test
statistic
print(f, p)
# Output :
```

```
1.5434275971616587 0.4523632544892888
```

Interpretation :

- Since p-value is >0.05, do not reject H0. There is no significant difference in variances of the two groups