Statistical Inference

Test for equality of variances

F-test for equality of variances

- The F test is used to test the equality of two population variances.
- Testing equality of variances is a prerequisite for many statistical tests (eg the 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 for the 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
    (The time taken to complete the MIS report should be normally distributed).

- Note: Generally the F test is used to the validate assumption of equal variance while performing the t test for equality of means. The parent population is assumed to follow a normal distribution.

Case Study - 1

Background

The company is analysing the time to complete an 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

Observations

time_g1	time_g2
85	83
95	85
105	96
85	94
90	102

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 an MIS report 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 second sample. n_1 and n_2 are sample sizes of the first and second sample respectively.**
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 R

```
# Import data
data<-read.csv("F test for 2 variances.csv",header=TRUE)

# Variance test
var.test(data$time_g1,data$time_g2,alternative = "two.sided")
```

- *time_g1,time_g2 are the variables under study.*
- *alternative="two.sided" , since under H1,variances are not equal.*

```
# Output :

        F test to compare two variances

data:  data$time_g1 and data$time_g2
F = 1.5434, num df = 11, denom df = 13, p-value = 0.4524
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4826988 5.2348866
sample estimates:
ratio of variances
          1.543428
```

Interpretation :

- *Since the p-value is >0.05, do not reject H0. There is no significant difference in variances of the two groups.*
- *Also, 95 percent confidence interval of ratio of variance contains 1,which means variances are same.*