**Equality of Variances in R-Homogeneity test**

Equality of Variances in R, in this article, we are describing variance comparison of 2 or more samples.

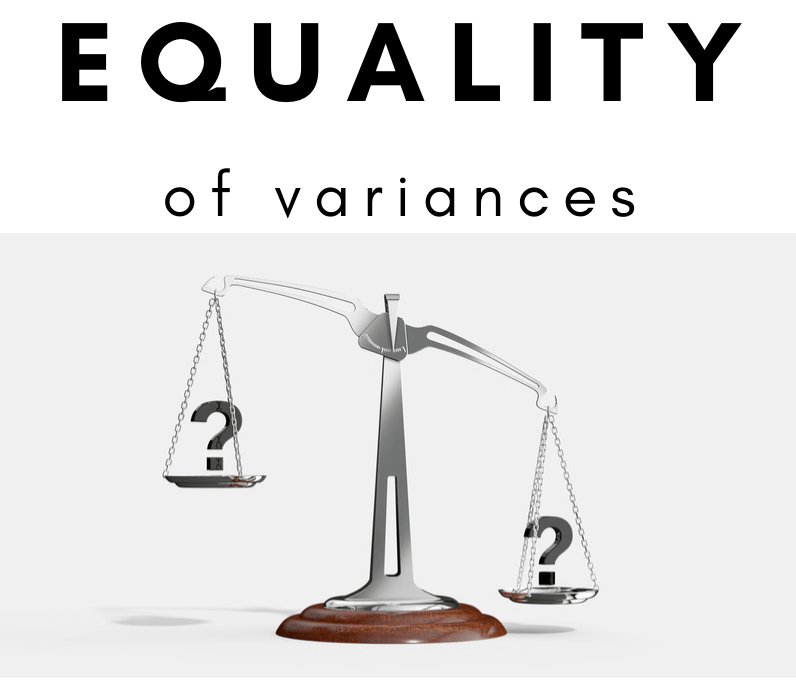
There are different types of tests that can be utilized to assess the equality of variances.

1) **F-test**:- Used for two groups variance comparison. Data must be normally distributed.

2) **Bartlett’s test**:- Used for two or more groups variance comparison. Data must be normally distributed.

3) **Levene’s test**:- An alternative to Bartlett’s test for non-normally distributed data.

4) **Fligner-Killeen’s test**:- A non-parametric test for non-normal data.



**Two sample equality of variances in R**

In most of the parametric methods, one of the assumptions is equality of the variances or some situation we want to measure the variability of the new instruments compared to old ones. Such kind of case F test will be very useful.

**1. F-test in R**

The F test statistic can be obtained by calculating the ratio of the two variances.

**F=VAR(A)/VAR(B)**

In the F test, the ratio deviates more from 1 then stronger the evidence of unequal variances.

Before doing the F test, we need to check one of the major assumptions is data should be normally distributed.

Normality distribution can be assessed based on the Shapiro test or visually using a QQ plot.

If any violation observed from normality then it better is to use Levene’s test or Fligner-Killeen test.

Levene’s test or Fligner-Killeen test is less sensitive and appropriate for when data is distributed non-normally.

Let’s see the F test syntax,

var.test(values ~ groups, data, alternative = "two.sided")

or

var.test(x, y, alternative = "two.sided")

ToothGrowth data set we used for F test calculation.

Let’s see the data structure,

str(ToothGrowth)

'data.frame':       60 obs. of  3 variables:

 $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...

 $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...

 $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...

Total 60 observations and 3 variables and the variable **supp** contains two groups. Let’s check the variance for the same.

Before that calculate the Shapiro test for the normality assumption validation,

**Shapiro Test:-**

shapiro.test(ToothGrowth$len)

Shapiro-Wilk normality test

data:  ToothGrowth$len

W = 0.96743, p-value = 0.1091

The p value is greater than 0.05, we can assume the normality.

res.ftest <- var.test(len ~ supp, data = ToothGrowth)

res.ftest

F test to compare two variances

data:  len by supp

F = 0.6386, num df = 29, denom df = 29, p-value = 0.2331

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.3039488 1.3416857

sample estimates:

ratio of variances

         0.6385951

The p-value of **0.2331** is greater than the significance level of 0.05. We can conclude that there is no significant difference between the two variances.

**Compare more than two sample variances in R**

When comparing more than two samples Bartlett’s test, Levene’s test, or Fligner-Killeen’s test will be more appropriate.

Coming to statistical hypotheses, Bartlett’s test, Levene’s test, or Fligner-Killeen’s test,

**Ho:** All populations variances are equal

**H1:** At least two of them different

When using Bartlett’s test one of the main assumptions data should be normally distributed. In the case of nonnormal data, the Levene test is an alternative to the Bartlett test.

If data is non-normally distributed, the Fligner-Killeen test is a non-parametric test alternative.

**2. Bartlett’s test in R**

Syntax of Bartlett’s test is

bartlett.test(formula, data)

Let’s check the equality of variance, We are using PlantGrowth dataset contains 30 observations and 2 variables.

str(PlantGrowth)

'data.frame':       30 obs. of  2 variables:

 $ weight: num  4.17 5.58 5.18 6.11 4.5 4.61 5.17 4.53 5.33 5.14 ...

 $ group : Factor w/ 3 levels "ctrl","trt1",..: 1 1 1 1 1 1 1 1 1 1 ...

The column group contains 3 factor variables ctrl, trt1, and trt2. Before doing Bartlett’s test let’s check the normality assumption.

**Shapiro Test:-**

shapiro.test(PlantGrowth$weight)

Shapiro-Wilk normality test

data:  PlantGrowth$weight

W = 0.98268, p-value = 0.8915

We can assume that data is normally distributed.

res <- bartlett.test(weight ~ group, data = PlantGrowth)

res

Bartlett test of homogeneity of variances

data:  weight by group

Bartlett's K-squared = 2.8786, df = 2, p-value = 0.2371

The p-value is  **0.2371** is greater than the significance level of 0.05. We can conclude that there is no significant difference between the tested sample variances.

**3. Levene’s test in R**

Levene test function is from car package, let’s load the library.

library(car)

leveneTest(weight ~ group, data = PlantGrowth)

Levene’s Test for Homogeneity of Variance (center = median)

      Df F value Pr(>F)

group  2  1.1192 0.3412

      27

The p-value is  **0.3412** is greater than the significance level of 0.05. We can conclude that there is no significant difference between the tested sample variances.

Levene’s test with multiple independent variables can check based on ToothGrowth dataset,

ToothGrowth dataset **dose** column stored as numeric variable let’s convert into factor variable first,

ToothGrowth$dose <- as.factor(ToothGrowth$dose)

leveneTest(len ~ supp\*dose, data = ToothGrowth)

Levene’s Test for Homogeneity of Variance (center = median)

      Df F value Pr(>F)

group  5  1.7086 0.1484

      54

**4. Fligner-Killeen test in R**

Will make use of the same data set,

fligner.test(weight ~ group, data = PlantGrowth)

Fligner-Killeen test of homogeneity of variances

data:  weight by group

Fligner-Killeen:med chi-squared = 2.3499, df = 2, p-value = 0.3088

The p-value is  **0.3088** is greater than the significance level of 0.05. We can conclude that there is no significant difference was observed between the tested sample variances.

**Summary**

This article provides one-of-a-kind exams for assessing the equality of variances in R among groups.