Statistical test for variance

F-test is used to assess whether the variances of two populations (A and B) are equal.

Uses:

Comparing two variances is useful in several cases, including:

- When you want to perform a two samples t-test to check the equality of the variances of the two samples
- When you want to compare the variability of a new measurement method to an old one. Does the new method reduce the variability of the measure?

Research questions and statistical hypotheses

Typical research questions are:

- whether the variance of group A (σ_A^2) is equal to the variance of group B (σ_b^2)?
- whether the variance of group A (σ_A^2 is less than the variance of group B (σ_A^2)?
- whether the variance of group A (σ_A^2) is greather than the variance of group B (σ_A^2)?

Hypothesis:

In statistics, we can define the corresponding null hypothesis (H_0) as follow:

- $H_0 : \sigma_A^2 = \sigma_b^2$
- $H_0 : \sigma_A^2 \leq \sigma_b^2$
- $H_0 : \sigma_A^2 \geq \sigma_b^2$

The corresponding alternative hypotheses (H_1) are as follow:

-$H_1 : \sigma_A^2 \neq \sigma_b^2$ (different)

-$H_1 : \sigma_A^2 > \sigma_b^2$ (greater)

-$H_1 : \sigma_A^2 < \sigma_b^2$ (less)

Requirement of F-test

Note that, the F-test requires the two samples to be normally distributed.

The R function *var.test()* can be used to compare two variances and its syntax is same as that of t- test.

two methods can be used:

Method 1

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

Method 2

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

Example

To illustrate F- test, I choose ToothGrowth data in R. A random sample of 10 individuals are shown bellow:

```
{r} # Store the data in the variable my_data my_data <- ToothGrowth library("dplyr")
sample_n(my_data, 10)
```

#Preleminary test to check F-test assumptions

F-test is very sensitive to departure from the normal assumption. You need to check whether the data is normally distributed before using the F-test.

Shapiro-Wilk test can be used to test whether the normal assumption holds. It's also possible to use Q-Q plot (quantile-quantile plot) to graphically evaluate the normality of a variable. Q-Q plot draws the correlation between a given sample and the normal distribution.

`{r} shapiro.test(my_data$len)` Since p- value is greater than 0.05. So the normality is holds good.

If there is doubt about normality, the better choice is to use *Levene's test* or *Fligner-Killeen test*, which are less sensitive to departure from normal assumption.

Calculating F- test

`{r} # F-test long method res.ftest <- var.test(len ~ supp, data = my_data, alternative="two.sided") res.ftest`

Interpretation of the result

The p-value of F-test is $p = 0.2331433$ which is greater than the significance level 0.05. In conclusion, there is no significant difference between the two variances.