WIKIPEDIA
The Free Encyclopedia

Contents

move to sidebar   hide

Levene's test

In statistics, **Levene's test** is an inferential statistic used to assess the equality of variances for a variable calculated for two or more groups.[1] Some common statistical procedures assume that variances of the populations from which different samples are drawn are equal. Levene's test assesses this assumption. It tests the null hypothesis that the population variances are equal (called *homogeneity of variance* or *homoscedasticity*). If the resulting p-value of Levene's test is less than some significance level (typically 0.05), the obtained differences in sample variances are unlikely to have occurred based on random sampling from a population with equal variances. Thus, the null hypothesis of equal variances is rejected and it is concluded that there is a difference between the variances in the population.

Some of the procedures typically assuming homoscedasticity, for which one can use Levene's tests, include analysis of variance and t-tests.

Levene's test is sometimes used before a comparison of means, informing the decision on whether to use a pooled t-test or the Welch's t-test. However, it was shown that such a two-step procedure may markedly inflate the type 1 error obtained with the t-tests and thus should not be done in the first place.[2] Instead, the choice of pooled or Welch's test should be made a priori based on the study design.

Levene's test may also be used as a main test for answering a stand-alone question of whether two sub-samples in a given population have equal or different variances.[3]

Levene's test was developed by and named after American statistician and geneticist Howard Levene.

Definition

Levene's test is equivalent to a 1-way between-groups analysis of variance (ANOVA) with the dependent variable being the absolute value of the difference between a score and the mean of the group to which the score belongs (shown below as $Z_{ij} = |Y_{ij} - \bar{Y}_{i \cdot}|$). The test statistic, W, is equivalent to the F statistic that would be produced by such an ANOVA, and is defined as follows:

$$W = \frac{(N-k)}{(k-1)} \cdot \frac{\sum_{i=1}^{k} N_i (Z_{i\cdot} - Z_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i\cdot})^2},$$

where

- k is the number of different groups to which the sampled cases belong,
- N_i is the number of cases in the ith group,
- N is the total number of cases in all groups,
- Y_{ij} is the value of the measured variable for the jth case from the ith group,
- $Z_{ij} = \begin{cases} |Y_{ij} - \bar{Y}_{i\cdot}|, & \bar{Y}_{i\cdot} \text{ is a mean of the } i\text{-th group,} \\ |Y_{ij} - \tilde{Y}_{i\cdot}|, & \tilde{Y}_{i\cdot} \text{ is a median of the } i\text{-th group.} \end{cases}$

(Both definitions are in use though the second one is, strictly speaking, the Brown–Forsythe test – see below for comparison.)

- $Z_{i\cdot} = \dfrac{1}{N_i} \sum_{j=1}^{N_i} Z_{ij}$ is the mean of the Z_{ij} for group i,

- $Z_{\cdot\cdot} = \dfrac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{N_i} Z_{ij}$ is the mean of all Z_{ij}.

The test statistic W is approximately F-distributed with $k-1$ and $N-k$ degrees of freedom, and hence is the significance of the outcome w of W tested against $F(1-\alpha; k-1, N-k)$ where F is a quantile of the F-distribution, with $k-1$ and $N-k$ degrees of freedom, and α is the chosen level of significance (usually 0.05 or 0.01).

Comparison with the Brown–Forsythe test

The Brown–Forsythe test uses the median instead of the mean in computing the spread within each group ($\bar{Y}$ vs. $\tilde{Y}$, above). Although the optimal choice depends on the underlying distribution, the definition based on the median is recommended as the choice that provides good robustness against many types of non-normal data while retaining good statistical power.[3] If one has knowledge of the underlying distribution of the data, this may indicate using one of the other choices. Brown and Forsythe performed Monte Carlo studies that indicated that using the trimmed mean performed best when the underlying data followed a Cauchy distribution (a heavy-tailed distribution) and the median performed best when the underlying data followed a chi-squared distribution with four degrees of freedom (a heavily skewed distribution). Using the mean provided the best power for symmetric, moderate-tailed, distributions.

See also

- Bartlett's test
- F-test of equality of variances

- Box's M test

References

1. Levene, Howard (1960). "Robust tests for equality of variances". In Ingram Olkin; Harold Hotelling; et al. (eds.). *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling.* Stanford University Press. pp. 278–292.

2. Zimmermann, Donald W. (2004). "A note on preliminary tests of equality of variances" (https://pub med.ncbi.nlm.nih.gov/15171807/). *British Journal of Mathematical and Statistical Psychology.* **57** (1): 173–81. doi:10.1348/000711004849222 (https://doi.org/10.1348%2F000711004849222).

3. Derrick, B; Ruck, A; Toher, D; White, P (2018). "Tests for equality of variances between two samples which contain both paired observations and independent observations" (http://jaqm.ro/iss ues/volume-13,issue-2/pdfs/3_BE_AN_DE_PA_.pdf) (PDF). *Journal of Applied Quantitative Methods.* **13** (2): 36–47.

External links

- Parametric and nonparametric Levene's test in SPSS (https://www.youtube.com/watch?v=O6taUl WejB0)
- http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm