Conditional vs. unconditional approaches to comparing two means with independent observations

Berenice López-Casal & Rocío Alcalá-Quintana[1]
Universidad Complutense de Madrid

Student's *t*-test is widely used to test the null hypothesis of equality of means with two independent samples of observations. It is a parametric technique that relies on a set of assumptions including normality (both samples have been drawn from normally distributed random variables) and homoscedasticity (both random variables have the same variance). Those assumptions can be put to test too, yielding a conditional strategy where statistical hypotheses about normality and homoscedasticity are tested first, and the choice of test of means depends on the outcome of test of assumptions. For instance, only if the null hypothesis about normality fails to be rejected on both groups a Student's *t*-test will be conducted; otherwise a robust alternative will be chosen. Likewise, if the null hypothesis about equality of variances is rejected, an alternative procedure like Welch's *t*-test (which does not assume homoscedasticity) will be performed over the means; otherwise a Student's *t*-test will be chosen. This conditional approach has gained popularity and it is customarily used even though it has been found to alter the type-I error rate of *t*-tests in such a way that the empirical rate no longer reflects the nominal significance level (Gans, 1981; Rochon & Kieser, 2011; Zimmerman, 1996, 2004).

This **app aims** at illustrating the effects of conditional and unconditional uses of *t*-tests on the empirical type-I error rate. The app simulates data and performs two-tailed tests for the difference of two means with independent observations in scenarios specified by the user.

Equal-variances assumption

This tab focuses on the homoscedasticity assumption. In the conditional procedure, homoscedasticity is tested first. When the null hypothesis about equality of variances is rejected, Welch's *t*-test is performed over the means; otherwise, a Student's *t*-test is performed. In the unconditional approach the two *t*-tests are always performed over the means without checking any assumptions.

Usage

To explore how violations of the equal-variances assumption affect empirical type-I error rates under conditional and unconditional approaches, samples are drawn from normally distributed variables with equal means (i.e., the null hypothesis for the *t*-test is always true).

The user can select the values of (i) the significance level (which corresponds to the nominal type-I error rate and is common to both, tests of variances and means), (ii) values of true standard deviations σ_1 and σ_2, and (iii) the sample sizes of each group. There are two tests of variances available: *F*-test and Levene's. Because the ratio of true standard deviations and the ratio of sample sizes affect

[1] To whom correspondence should be addressed (ralcala@psi.ucm.es)

performance of *t*-tests, they are explicitly shown on the sidebar. In addition, true distributions of the variables are plotted at the bottom of the sidebar, alongside a legend with the values of the standard deviations σ, skewness γ, and excess kurtosis κ (group #1 to the left and colored red, group #2 to the right and colored green). Since only normal distributions are used, $\kappa_1 = \kappa_2 = 0$, and $\gamma_1 = \gamma_1 = 0$.

Output

Results for the conditional and unconditional approaches are respectively shown on the left and right columns. The **top-left histogram** plots the empirical distribution of *p*-values yielded by Student's *t*-test (red) and Welch's *t*-test (blue) used as prescribed by the conditional procedure. Numeric labels on the top left area correspond to the overall proportions of null hypotheses rejected (type-I error rate, left) and not rejected (right). There is a drop-down menu above the histogram that allows users to display only the *p*-values of Student's test, only those of Welch's test, or both. If only one type of test is selected, light bars showing results for that test under the unconditional approach are included for reference, and the numeric labels on the top left represent proportions for that test alone.

The **histogram at the bottom** (colored green) shows the distribution of *p*-values of the test of equality of variances (either *F* or Levene's, depending on the user's choice), alongside its null-hypothesis rejection and non-rejection rates.

The two **histograms on the right** depict the empirical distribution of *p*-values and empirical type-I error rate for Student's *t*-test (top) and Welch's *t*-test (bottom) without following a conditional approach (i.e., all samples are tested using both tests).

Additional plots can be accessed by clicking the button between the information and warning icons, which show empirical type-I error rates for *t*-tests as a function of the ratio between standard deviations and the ratio between sample sizes. Further explanation is included in the figure caption.

Normality assumption

This tab focuses on the normality assumption. In the conditional procedure, a goodness-of fit test on normality is carried out first and only when the null hypothesis fails to be rejected in both groups a Student's *t*-test is conducted. When normality is rejected in at least one group, Cressie and Whitford's (1986) robust U_2 *t*-test is used as a replacement.

Usage

To explore how violations of the normality assumption affect empirical type-I error rates under conditional and unconditional approaches, samples are drawn from variables with equal variances (i.e., the homoscedasticity assumption is met) along with equal means (i.e., the null hypothesis for the *t*-test is always true). The user can select the significance level, α (nominal type-I error rate), the shape of the distributions, and the sample sizes n_1 and n_2 of each group.

Along with the significance level, an option to correct for multiple comparisons is offered. Because Student's *t*-test assumes normality for both groups, two separate tests of normality must be carried out per test on means. In order to maintain an overall significance level of α, a cutoff point of α/2 must be used for each normality test.

Choices of shape are (i) normal (symmetrical and mesokurtic), (ii) uniform for a symmetric and platykurtic distribution, (iii) Laplace for a symmetric and leptokurtic option, (iv) lognormal for a strongly positively skewed variable, and (v) beta with parameters chosen to render a moderately negatively skewed alternative. True distributions selected by the user are plotted at the bottom of the sidebar, alongside a legend with the values of the standard deviations σ, skewness γ, and excess kurtosis κ (group #1 to the left and colored red, group #2 to the right and colored green).

Output

Results for the conditional and unconditional approaches are respectively shown on the left and right columns. The **top-left histogram** depicts the empirical distribution of *p*-values yielded by Student's *t*-test (red) and Cressie's *t*-test (grey), used as prescribed by the conditional procedure. Numeric labels over the arrows in the top left area correspond to the proportions of null hypotheses on means rejected (type-I error rate, left) and not rejected (right) under the conditional approach. There is a drop-down menu above the histogram that allows users to display only the *p*-values of Student's test, only those of Cressie's *t*-test, or both. If only one type of test is selected, light bars showing results for that test under the unconditional approach are included for reference, and the numeric labels on the top left represent proportions for that test alone.

The **histogram at the bottom** (green) shows the distribution of *p*-values of the goodness-of-fit test of normality (either, Shapiro-Wilk or chi-square depending on the user's input) for groups #1 (45° patterned bars, bottom) and #2 (-45° patterned bars, top). Individual rejection rates are shown separately (top right legend) for both groups whereas the combined rejection and non-rejection rates are printed above the arrows. Only combined rejection rates are used in the conditional approach, as it requires rejection of normality in at least one group.

The **histograms on the right** depict the empirical distribution of *p*-values and empirical type-I error rate for Student's (top) and Cressie's (bottom) *t*-tests under the unconditional approach (i.e., performed over all samples).

Additional plots can be accessed by clicking the button between the information and warning icons, which show empirical type-I error rates for *t*-tests as a function of the ratio between sample sizes when one of the samples has been drawn from a non-normal distribution (uniform, Laplace, lognormal, and beta) and the other from a normal one. Further explanation is available in the figure caption.

Equal-variances and Normality assumptions

This tab allows examining the homoscedasticity and normality assumptions in a three-step procedure: First of all, a test on normality is performed. If normality is rejected for at least one group, Cressie & Whitford's (1986) modification of their U_2 test, which assumes neither normality nor homoscedasticity, is conducted. Otherwise, variances are tested leading to either a Welch's *t*-test on means (upon rejection of homoscedasticity) or a Student's *t*-test on means (if homoscedasticity cannot be rejected).

Usage

As in previous tabs, samples are drawn from variables with equal means (i.e., the null hypothesis for the *t*-test is always true). There are two goodness-of-fit tests available (Shapiro-Wilk and chi-square) and two tests of variances (*F* and Levene's). The user can also select (i) the significance level, α (which is used in all tests), (ii) whether a correction for multiple comparisons is applied to tests on normality, (iii) the shape of the distributions from which samples are drawn, (iv) the values of true standard deviations σ_1 and σ_2, and (v) the sample sizes n_1 and n_2 of each group.

Because the ratio of standard deviations and the ratio of sample sizes affect the performance of *t*-tests, they are explicitly shown on the sidebar. Distributions selected by the user are plotted at the bottom of the sidebar, alongside a legend with the values of the standard deviations σ, skewness γ, and excess kurtosis κ (group #1 to the left and colored red, group #2 to the right and colored green).

Output

Results for the conditional and unconditional approaches are respectively shown on the left and right columns. The **histogram on the left** depicts the empirical distribution of p-values yielded by Student's *t*-test (red), Welch's *t*-test (light blue), and Cressie's *t*-test (dark blue) used as prescribed by the conditional procedure. Numeric labels over the arrows in the top left area correspond to the overall proportions of null hypotheses on means rejected (type-I error rate, left) and not rejected (right) under the conditional approach. There is a drop-down menu above the histogram that allows users to display only the *p*-values of Student's test, only those of Welch's *t*-test, those of Cressie's *t*-test, or all three. If only one type of test is selected, light bars showing results for that test under the unconditional approach are included for reference, and the numeric labels on the top left represent proportions for that test alone.

The two **histograms on the right** (unconditional) depict the empirical distribution of *p*-values and empirical type-I error rate for Student's *t*-test (top), Welch's *t*-test (middle), and Cressie's *t*-test (bottom) without following a conditional approach (i.e., all tests are performed on all samples without any prior checks).

Plots showing the distribution of *p*-values for the tests of normality and homoscedasticity are available by clicking the button between the information and warning icons. Further explanation is available in the figure caption.

References

Cressie, N. A. C. & Whitford, H. J. (1986) How to use the two-sample *t*-test, *Biometrical Journal*, *2*, 131–148.

Gans, D. J. (1981). Use of a preliminary test in comparing two sample means. *Communications in Statistics - Simulation and Computation*, *10*, 163-174.

Rochon, J., & Kieser, M. (2011). A closer look at the effect of preliminary goodness-of-fit testing for normality for the one-sample *t*-test. *British Journal of Mathematical and Statistical Psychology, 64*, 410-426.

Zimmerman, D. W. (1996). Some properties of preliminary tests of equality of variances in the two-sample location problem. *The Journal of General Psychology, 123*, 217-231.

Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology, 57*, 173-181.