Homogeneity of Variance in the Two-Sample Means Test

Barry K. Moser; Gary R. Stevens

The American Statistician, Vol. 46, No. 1 (Feb., 1992), 19-21.

Stable URL:
http://links.jstor.org/sici?sici=0003-1305%28199202%2946%3A1%3C19%3AHOVITT%3E2.0.CO%3B2-%23

The American Statistician is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/astata.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Sat Aug 27 17:43:57 2005

In this department *The American Statistician* publishes articles, reviews, and notes of interest to teachers of the first mathematical statistics course and of applied statistics courses. The department includes the Accent on Teaching Materials section; suitable contents for the section are described under the section heading. Articles and notes for the department, but not intended specifically for the section, should be useful to a substantial number of teachers of the indicated types of courses or should have the potential for fundamentally affecting the way in which a course is taught.

Homogeneity of Variance in the Two-Sample Means Test

BARRY K. MOSER and GARY R. STEVENS*

In the two-sample means test, the need for a preliminary variance test and the special emphasis given to the equal variance assumption are questioned. The conclusions can be easily used in the classroom to teach the two sample means test.

KEY WORDS: Size; Power; Preliminary variance test.

The problem of testing the equality of the means from two independent normally distributed populations is covered in every elementary statistics textbook. Under the assumption that the two population variances σ_1^2 and σ_2^2 are equal, most authors recommend the t test. If the variances are unequal, then an alternative procedure, such as the method suggested by Smith (1936), by Welch (1937) and later by Satterthwaite (1946), is often recommended. To apply such an approach, the variance ratio, $\theta = \sigma_2^2/\sigma_1^2$, must be known. In most applications, however, the variance ratio is unknown. In such cases, it is current practice to apply a formal preliminary test for $\sigma_1^2 = \sigma_2^2$. If $\sigma_1^2 = \sigma_2^2$ is accepted in the preliminary test, then a t test is performed on the means; otherwise a Smith/Welch/Satterthwaite (SWS) test on the means is performed.

Preliminary variance tests have been incorporated into the two-sample means tests of popular statistical packages, such as SAS, BMDP and SPSS. All of these packages have two-sample means tests that include a preliminary equality of variance test. The preliminary variance test is calculated first, followed by the choice between a t test and an SWS test. To illustrate how the statistical routines are used, we examine an example presented by Milton and Arnold (1986, p. 312, example 10.6.1). In the example, two brands of kerosene heaters are tested. The observations are the times required to raise the room temperature 10°F. We analyzed the logarithm of the time using SAS PROC TTEST. The SAS output is listed in Table 1. For each brand of heater, the procedure lists the sample size, means, standard deviations, standard errors, minimum and maximum values. A preliminary equality of variance test follows, where the calculated F ratio, the degrees of freedom, and the P value are presented. The SWS and t statistics are listed in the rows titled Unequal and Equal Variances, respectively. Thus, the user checks the equality of variance test, if equality is accepted, he proceeds to the t statistic, if equality is not accepted, he uses the SWS statistic.

It is apparent from this discussion that a great deal of emphasis is placed on the homogeneity of variance assumption in the two-sample means test. Authors of elementary statistics textbooks stress the importance of the assumption, to the extent that textbooks are divided into sections where the t test is presented if variance equality is assumed and the SWS test is presented if variance equality is violated. Likewise, popular statistical packages have incorporated preliminary variance tests in their two-sample means test routines. The following questions arise: Is the current practice of preliminary variance tests appropriate and is this emphasis on variance homogeneity warranted?

The objective of this article is to address these questions. To this end, the relationship between the t test and SWS test is examined when the choice between the two procedures is based on a preliminary variance test. In the following sections, the problem is outlined, the sizes and powers of the means tests are summarized, and the two questions are answered.

THE PROBLEM

Let $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ be independent random samples from two normally distributed populations, where $x_i \sim N(\mu_1, \sigma_1^2)$ and $y_j \sim N(\mu_2, \sigma_2^2)$, for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$.

The preliminary test for H_0: $\sigma_1^2 = \sigma_2^2$ versus H_1: $\sigma_1^2 \neq \sigma_2^2$ is to calculate

$$F' = s_2^2/s_1^2 \tag{1}$$

and reject H_0 if $F' > F_{n_2-1,n_1-1}^{\alpha/2}$ or $F' < F_{n_2-1,n_1-1}^{1-\alpha/2}$, where α is the prescribed variance test significance level, $F_{a,b}^{\alpha^*}$

*Barry K. Moser is Associate Professor, Department of Statistics, Oklahoma State University, Stillwater, 74078. Gary R. Stevens is Senior Statistician, Sterling Drug, 81 Columbia Turnpike, Rensselaer, NY 12144.

© 1992 American Statistical Association

Table 1. SAS PROC TTEST Output for Kerosene Heater Example

TTEST PROCEDURE										
Brand	N	Mean	Standard deviation	Standard error	Min	Max	Variances	T	DF	Prob > |T|
B	12	3.685	.487	.141	2.62	4.24	Unequal	1.93	16.4	.0710
A	15	3.381	.273	.071	2.63	3.73	Equal	2.05	25.0	.0508

NOTE: For H_0: Variances are equal, $F' = 3.18$ DF = (11, 14) Prob > F' = .0449. Dependent variable = logarithm of time in seconds.

is the $100(1 - \alpha^*)$ percentile point of an F distribution with a and b degrees of freedom, $s_1^2 = (n_1 - 1)^{-1} \Sigma_{i=1}^{n_1} (x_i - \bar{x})^2$, $s_2^2 = (n_2 - 1)^{-1} \Sigma_{j=1}^{n_2} (y_j - \bar{y})^2$, $\bar{x} = \Sigma_{i=1}^{n_1} x_i/n_1$, and $\bar{y} = \Sigma_{j=1}^{n_2} y_j/n_2$. The above rejection criterion is equivalent to the P value criterion that $\sigma_1^2 = \sigma_2^2$ is rejected for all P values less than α.

If H_0 is not rejected, then the t test for H_0^*: $\mu_1 = \mu_2$ versus H_1^*: $\mu_1 \neq \mu_2$ (or H_0^*: $\mu_1 = \mu_2$ versus H_2^*: $\mu_1 > \mu_2$) is to calculate

$$t^* = \frac{(\bar{x} - \bar{y})}{s_p \sqrt{(1/n_1 + 1/n_2)}} \tag{2}$$

and reject H_0^* if $t^{*2} > F^{\delta}_{1,n_1+n_2-2}$ (or reject H_0^* if $t^* > t_{\delta,n_1+n_2-2}$), where δ is the prescribed means test significance level and $s_p^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2)$.

If H_0 is rejected, then the SWS criterion for testing H_0^* versus H_1^* (or H_0^* versus H_2^*) is to calculate

$$t^{**} = \frac{(\bar{x} - \bar{y})}{\sqrt{(s_1^2/n_1 + s_2^2/n_2)}} \tag{3}$$

and reject H_0^* if $t^{**2} > F^{\delta}_{1,\nu}$ (or reject if H_0^* if $t^{**} > t_{\delta,\nu}$), for the same prescribed significance level δ given in (2), with

$$\nu = \frac{(1/n_1 + u/n_2)^2}{1/n_1^2(n_1 - 1) + u^2/n_2^2(n_2 - 1)} \tag{4}$$

and $u = s_2^2/s_1^2$.

It should be noted that, if the prescribed variance test significance level, α, is set to 0, then H_0 is accepted and the t test (2) for H_0^* is always performed. Likewise, if α is set to 1, then the SWS test (3) for H_0^* is always performed. Therefore, specifying an α level of 0 (or 1) is equivalent to performing a t test (or SWS test) without any preliminary test for H_0. For purposes of identification, the preliminary test for H_0 with $0 < \alpha < 1$ followed by a t test or SWS test for H_0^* will be referred to as the Sometimes t (ST) test. Performing a t test for H_0^* without any preliminary test for H_0 (i.e., $\alpha = 0$), will be referred to as an Always t (AT) test. Using a SWS test for H_0^* without any preliminary test for H_0 (i.e., $\alpha = 1$) will be referred to as an Always SWS (ASWS) test.

THE SIZE AND POWER OF THE TEST

The probability of rejecting H_0^*: $\mu_1 = \mu_2$, in favor of H_1^* or H_2^*, is a function of n_1, n_2, θ, δ, α, and $\lambda = (\mu_2 - \mu_1)^2/[2(\sigma_1^2/n_1 + \sigma_2^2/n_2)]$. When $\lambda = 0$, the probability of rejecting H_0^* is the size of the test. The power of the test corresponds to the probability of rejecting H_0^* for any $\lambda > 0$. In either case, the probability of rejecting H_0^* is expressed as the sum of two mutually exclusive alternatives:

$$P_r(\text{reject } H_0^*) = P_r(\text{reject } H_0^* \text{ and do not reject } H_0) + P_r(\text{reject } H_0^* \text{ and reject } H_0). \tag{5}$$

Sizes and powers for combinations of n_1, n_2, θ, δ, α, and λ were calculated by Moser, Stevens, and Watts (1989). Their results are summarized as:

(a) If the sample sizes are equal, then the ASWS, ST and AT tests have almost identical sizes and powers for all n_1, n_2, $\lambda > 0$, and $1 \leq \theta \leq 10$.

(b) If the sample sizes are unequal but the variance ratio, θ, is near one, then the AT test attains the largest power while maintaining a size near the prescribed δ level.

(c) If the sample sizes are unequal, θ is not near one, and the smaller variance is associated with the larger sample size, then the AT test still has the largest power for all $\lambda > 0$, followed by the ST and ASWS tests. However, the AT test attains this large power at the cost of a large size. For certain values of θ, the ST test will also attain its larger power at the cost of a large size. The ASWS test, however, maintains a reasonable size near δ for all $1 \leq \theta \leq 10$.

(d) If the sample sizes are unequal, θ is not near one, and the smaller variance is associated with the smaller sample size, then the ASWS test has the largest power, followed by the ST and AT tests. The ASWS test continues to maintain a reasonable size near δ for all $1 \leq \theta \leq 10$.

DISCUSSION AND RECOMMENDATIONS

We now address the question: Is the current practice of preliminary variance tests appropriate? The answer is no. We justify this response in the following paragraphs.

The t test provides the largest power while maintaining a size near the prescribed δ level whenever the sample sizes are unequal and the variance ratio is near one. Therefore, if the sample sizes are unequal and the variance ratio is *known* to be near one, then the t test is appropriate.

The SWS, the ST and the t tests all have the same sizes and powers whenever the sample sizes are equal. Therefore, a preliminary variance test is superfluous, as it simply adds an extra analysis step. Thus, in the equal sample size case, either the t test or the SWS test is appropriate.

The only remaining situations occur when the sample sizes are unequal and the variance ratio is unknown or known to differ from one. Whenever the sample sizes are unequal, the SWS test provides good power while maintaining a reasonable size near δ. However, both the ST and the t test can produce large sizes when the variance ratio differs from one. Therefore, the SWS test is appropriate here.

Note that the ST test with its preliminary variance test was never recommended. For all values of the variance ratio, no sample size combinations exist where both the t test and the SWS test perform poorly while the ST test performs well. This is due to the fact that the size (power) of the ST test is a weighted average of the sizes (powers) of the t test and the SWS test. Therefore, preliminary variance test (1) is never appropriate for this problem.

Finally, is the emphasis on variance homogeneity warranted? Again the answer is no. The variance ratio influences the choice of the appropriate test procedure only when the variance ratio is known. In practice, however, the variance ratio is rarely known. From the above discussion, the SWS test is appropriate in all cases where the variance ratio is unknown. Therefore, when teaching the two-sample means test, more effort should be spent learning the qualities of the SWS test and less emphasis should be placed on the homogeneity of variance assumption.

[*Received November 1989. Revised January 1991.*]

REFERENCES

Milton, J. S., and Arnold, J. C. (1986), *Probability and Statistics in the Engineering and Computing Sciences,* New York: McGraw-Hill.

Moser, B. K., Stevens, G. R., and Watts, C. L. (1989), "The Two-Sample t Test Versus Satterthwaite's Approximate F Test," *Communications in Statistics—Theory and Methods,* 18.

Satterthwaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin,* 2, 110–114.

Smith, H. F. (1936), "The Problem of Comparing the Results of Two Experiments With Unequal Errors," *Journal of the Council for Scientific and Industrial Research,* 9, 211–212.

Welch, B. L. (1937), "The Significance of the Difference Between Two Means When the Population Variances are Unequal," *Biometrika,* 29, 350–362.