O'Brien Test for Homogeneity of Variance

Hervé Abdi[1]

1 Overview

The homogeneity of variance assumption is one of the critical assumptions underlying most parametric statistical procedures such as the analysis of variance and it is important to be able to test this assumption. In addition, showing that several samples do not come from populations with the same variance is sometimes of importance *per se*. Among the many procedures used to test this assumption, one of the most sensitive is the *O'Brien* test. This test was developed by O'Brien (1979, 1981). The null hypothesis for this test is that the samples under considerations come from populations with the same variance; The alternative hypothesis is that the populations have different variances.

Compared to other tests of homogeneity of variance, the advantage of the O'Brien test resides in its versatility and its compatibility with standard analysis of variance designs (*cf.* O'Brien, 1979; Martin & Games, 1977; Games, Keselman, & Clinch 1979). It is also optimal because it minimizes both Types I and II errors. The

[1]In: Neil Salkind (Ed.) (2007). *Encyclopedia of Measurement and Statistics.* Thousand Oaks (CA): Sage.
Address correspondence to: Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/~herve

essential idea behind the O'Brien test is to replace, for each sample, the original scores by transformed scores such that the transformed scores reflect the variance of the sample. Then, a standard analysis of variance based on the transformed scores will test the homogeneity of variance assumption.

1.1 Motivation and method

There are several tests available to detect if several samples come from populations having the same variances. In the case of two samples, the ratio of the population estimates (computed from the samples) is distributed as a Fisher distribution under the usual assumptions. Unfortunately there is no straightforward extension to this approach for designs involving more than two samples. By contrast, the O'Brien test is designed to test the homogeneity of variance assumption for several samples at once and with the versatility for analysis of variance designs including contrast analysis and sub-designs analysis.

The main idea behind the O'Brien test is to transform the original scores so that the transformed scores reflect the variation of the original scores. An analysis of variance on the transformed scores will then reveal differences in the variability (*i.e.,* variance) of the original scores and therefore this analysis will test the homogeneity of variance assumption. A straightforward application of this idea will be to replace the original scores by the absolute value of their deviation to the mean of their experimental group (*cf.* Levene, 1960; Glass & Stanley, 1970). So, if we denote by Y_{as} the score of subject s in experimental condition a whose mean is denoted by $M_{a.}$, this first idea amounts to transforming Y_{as} into v_{as} as:

$$v_{as} = \left| Y_{as} - M_{a.} \right| .$$

This transformation has the advantage of being simple and easy to understand, but, unfortunately, it creates some statistical problems (*i.e.,* the F distribution does not model the probability distribution under the null hypothesis) and in particular, it leads to an excess of Type I errors (*i.e.,* we reject the null hypothesis more of-ten than the α level indicates, see, *e.g.*,, Miller, 1968; Games *et al.*, 1977; O'Brien, 1979 for more details).

A better approach is to replace each score by its absolute distance to the *median* of its group. Specifically, each score is replaced by

$$w_{as} = \left| Y_{as} - Md_{a.} \right|$$

with $Md_{a.}$: median of Group a. This transform gives very satisfactory results for an omnibus testing of the homogeneity of variance assumption. However in order to implement more sophisticated statistical procedures (*e.g.,* contrast analyses , multiples comparisons), a better transformation has been proposed by O'Brien (1979, 1981). Here, the scores are transformed as:

$$u_{as} = \frac{N_a (N_a - 1.5)(Y_{as} - M_{a.})^2 - .5 SS_a}{(N_a - 1)(N_a - 2)} ,$$

with:

- N_a Number of observations of Group a
- $M_{a.}$ Mean of Group a
- SS_a Sum of the squares of Group a: $SS_a = \sum_s (Y_{as} - M_{a.})^2$.

When all the experimental groups have the same size, this formula can be simplified as

$$u_{as} = \frac{N (N - 1.5)(Y_{as} - M_{a.})^2 - .5 SS_a}{(N - 1)(N - 2)}$$

(with N: number of observations per group).

2 Example

In this section we detail the computation of the Median and the O'Brien transforms. We use data from a memory experiment reported by Hunter (1964, see also Abdi, 1987).

2.1 One is a bun ...

In this experiment, Hunter wanted to demonstrate that it is easier to remember an arbitrary list of words when we use a mnemonic devise such as the peg-word technique. In this experiment, 64 participants were assigned to either the control or the experimental group. The task for all participants was to learn an arbitrary list of pairs of words such as "one-sugar," "two-tiger," ... "ten-butterfly." Ten-minute after they had learned their list, the participants were asked to recall as many pairs as they could. Participants in the control group were told to try to remember the words as best as they could. Participants from the experimental group were given the following instructions:

> A good way to remember a list is to first learn a "nursery-rhyme" such as: *"one is a bun, two is a shoe, three is a tree, four is door, five is a hive, six is a stick, seven is heaven, eight is a gate, nine is a mine, and ten is a hen."* When you need to learn a pair of words, start by making a mental image of the number and then make a mental image of the second word and try to link these two images. For example, in order to learn "one-cigarette" imagine a cartoon-like bun smoking a cigarette.

The results are given in Table 1.

The results of this experiment are illustrated in Figure 1, they show that the participants from the experimental group are doing better than the participants from the control group. To confirm this interpretation, an analysis of variance was performed (see Table 2) and the F-test indicates that, indeed, the average number of words recalled is significantly larger in the experimental group than in the control group.

Figure 1 also shows that a large proportion of the participants of the experimental are getting a perfect score of 10 out of 10 words (*cf.* the peak at 10 for this group). This is called a *ceiling* effect: Some of the participants of the experimental group could have performed even better if they had had more words to learn. As a consequence of this ceiling effect, the variance of the experimental group is likely to be *smaller* than what it should be because the

Table 1: Data from Hunter (1964). The table gives for each group the frequency of subjects recalling a given number of words. For example, 11 subjects in the control group recalled 6 words from the list they had learned.

Number of Words Recalled	Control Group	Experimental Group
5	5	0
6	11	1
7	9	2
8	3	4
9	2	9
10	2	16
$Y_{a.}$	216	293
$M_{a.}$	6.750	9.156
$Md_{a.}$	6.500	9.500
SS_a	58.000	36.219

Table 2: ANOVA Table for the experiment of Hunter (1964). Raw data.

Source	df	SS	MS	F	Pr(F)
Experimental	1	92.64	92.64	60.96**	.000 000 001
Error	62	94.22	1.52		
Total	63	186.86			

**: p smaller than $\alpha = .01$. $R^2_{\mathscr{A}.Y} = .496$.

ceiling effect eliminates the differences between the participants with a perfect score. In order to decide if this ceiling effect does

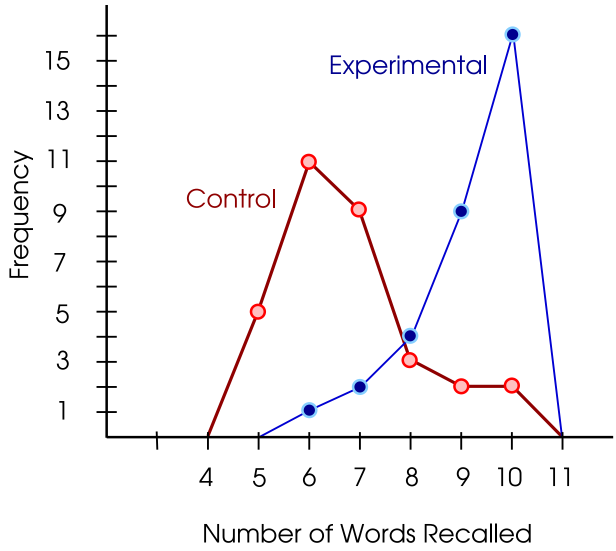

Figure 1: Results of the "peg-word" experiment (one is a bun) from Hunter (1964).

reduce the size of the variance of the experimental group, we need to compare the variance of the two groups as shown below.

The first step to test the homogeneity of variance is to transform the original scores. For example, the transformation of Y_{as} into w_{as} for Observation number 5 from the control group gives

$$w_{as} = \left| Y_{as} - Md_{a.} \right| = 5 - 6.5 = 1.5 \;.$$

The transformation of Y_{as} into u_{as} for Observation number 5 from the control group gives

$$\begin{aligned} u_{as} &= \frac{N\,(N-1.5)\,(Y_{as}-M_{a.})^2 - .5SS_a}{(N-1)\,(N-2)} \\ &= \frac{32\,(32-1.5)\,(5-6.75)^2 - .5\times 58}{31\times 30} \\ &= 3.1828 \;. \end{aligned}$$

The recoded scores are given in Tables 3 and 4. The ANOVA table obtained from the analysis of transformation w_{as} is given in

Table 3: Recoded scores: control group.

	control group		
Number of words recalled	Frequency	w_{as}	u_{as}
5	5	1.5	3.1828
6	11	0.5	0.5591
7	9	0.5	0.0344
8	3	1.5	1.6086
9	2	2.5	5.2817
10	2	3.5	11.0538

Table 4: Recoded scores: experimental group.

	Experimental Group		
Number of words recalled	Frequency	w_{as}	u_{as}
5	0	4.5	—
6	1	3.5	10.4352
7	2	2.5	4.8599
8	4	1.5	1.3836
9	9	0.5	0.0061
10	16	0.5	0.7277

Table 5: ANOVA.Homogeneity of variance test. Recoded scores: w_{as}. (Median)

Source	*df*	*SS*	*MS*	*F*	Pr(*F*)
Experimental	1	0.77	0.77	1.16^{ns}	.2857
Error	62	41.09	0.66		
Total	63	41.86			

ns: No significant difference.

Table 6: ANOVA.Homogeneity of variance test. Recoded scores: u_{as}. (O'Brien test).

Source	*df*	*SS*	*MS*	*F*	Pr(*F*)
Experimental	1	7.90	7.90	1.29^{ns}	.2595
Error	62	378.59	6.11		
Total	63	386.49			

ns: No significant difference..

Table 5. The ANOVA table obtained from the analysis of transformation u_{as} is given in Table 6.

Looking at Tables 5 and 6 indicates that we cannot show that the ceiling effect observed in Figure 1 significantly reduces the variance of the experimental group compared to the control group.

References

[1] Abdi, H. (1987). *Introduction au traitement statistique des données expérimentales.* Grenoble: Presses Universitaires de Grenoble.

[2] Games, P.A., Keselman, H.J., & Clinch, J.J. (1979). Tests for homogeneity of variance in factorial designs. *Psychological Bulletin,* **86**, 978–984.

[3] Glass, G.V., & Stanley, J.C. (1970). *Statistical methods in education and psychology.* Englewood Cliffs: Prentice-Hall.

[4] Hunter, I.M.L. (1964). *Memory.* London: Penguin.

[5] Levene, H. (1960). Robust test for the equality of variances. In Olkin, I. (Ed.), *Contributions to probability and statistics.* Palo Alto: Standford University Press.

[6] Martin, C.G., & Games, P.A. (1977). ANOVA tests for homogeneity of variance: Non-normality and unequal samples *Journal of Educational statistics,* **2**, 187–206

[7] Miller, R.G. (1968). Jackkniffiing variances. *Annals of Mathematical Statistics,* **39**, 567–582.

[8] O'Brien, R.G. (1979). A general ANOVA method for robust test of additive models for variance. *Journal of the American Statistical Association,* **74**, 877–880.

[9] O'Brien, R.G. (1981). A simple test for variance effects in experimental designs. *Psychological Bulletin,* **89**, 570–574.