Williams' Test Comparing Multiple Dose Groups to a Control

Morton B. Brown and Shibao Feng
University of Michigan, Department of Biostatistics
1420 Washington Heights
Ann Arbor, MI 48109-2029, USA
mbb@umich.edu

1. Introduction

In dose-ranging pharmaceutical trials, several groups of subjects who receive active doses are compared to a group receiving a placebo or conventional therapy. When the effect of the dose on the mean level of the response is expected to be monotonic, Williams (1971, 1972) proposed a test to compare multiple groups to a control. He also computed tables of the critical values for the simple design that corresponds to a one-way analysis of variance (ANOVA). Guo and Brown (1996) compared the power of Williams' test with those of the step-down procedure and Dunnett's multiple comparison method. Williams' test had acceptable power over all orderings of the means, but was substantially more powerful than its two competitors for some orderings of the observed effects.

In this paper we consider the implementation of Williams' test in a more general setting, that of the general linear model. Even when the primary comparison is between groups, it is necessary to adjust for potential confounding effects, such as centre, age or gender. In section 2 we show that the 5% critical values of Williams' test correspond to the 4^{th} percentile of Student's t-distribution. More extreme critical values correspond to a percentile of Student's t that is less than 0.8 of the desired size. Based on this we propose that the usual p-values obtained by contrasts that are used in Williams' test be multiplied by 1.25 to approximate the p-values for Williams' test. Using simulation, we evaluate the empirical sizes of the proposed approximation to show that the approximation is conservative.

2. The size of Williams' critical values using the Student t-distribution

Table 1 presents tail probabilities of the Student t-distribution if they are evaluated using Williams' critical values. When the critical values are multiplied by 1.25, they will be equal to (subject to rounding error) or exceed the nominal size of Williams' test. Therefore, the use of 1.25 to rescale p-values obtained for testing a contrast corresponding to Williams' test should be conservative.

2. The empirical size of Williams' test using a GLM procedure

To perform Williams' test using a general linear model procedure, it is necessary to specify a set of contrasts corresponding to the effects that are tested. Consider a design consisting of three dose levels and a placebo. Let μ_0, μ_1, μ_2, μ_3 represent the means of the placebo group and the three dose levels in ascending order. Assuming equal sample sizes, the three contrasts that need to be computed are estimates of $\mu_3 - \mu_0$, $(\mu_3 + \mu_2)/2 - \mu_0$, $(\mu_3 + \mu_2 + \mu_1)/3 - \mu_0$. The largest observed difference is used in the test. It is divided by an estimate of scale to enter the tables for Williams' test; as the estimate of scale Williams used the square root of the pooled variance multiplied by *2/r* where *r* is the sample size for each group.

Table 1. Tail probabilities of Student's t-distribution using critical values of Williams' test

No. of dose levels	DF	Williams upper 5% critical value	p-value based on Student t-distribution	Williams upper 2.5% critical value	p-value based on Student t-distribution	Williams upper 1% critical value	p-value based on Student t-distribution
2	20	1.81	0.0427	2.155	0.0218	2.58	0.0089
	40	1.76	0.0430	2.083	0.0218	2.47	0.0089
	120	1.73	0.0431	2.037	0.0219	2.40	0.0090
3	20	1.83	0.0411	2.177	0.0208	2.60	0.0086
	40	1.79	0.0405	2.102	0.0209	2.48	0.0087
	120	1.75	0.0413	2.055	0.0210	2.41	0.0087
4	20	1.84	0.0397	2.187	0.0204	2.61	0.0084
	40	1.80	0.0397	2.111	0.0205	2.49	0.0085
	120	1.77	0.0396	2.063	0.0206	2.42	0.0085

A simulation was conducted with one covariate where the response variable y was generated from the formula $y = \beta Z_1 + Z_2$ where β was determined by a specified correlation between y and Z_1 (=0,0.3,0.5,0.7,0.9). The simulations were performed for group sample sizes equal to 20, 40 and 120. A test was significant only if it had a p-value less than 0.04 using Student's t-distribution. The results of the simulation are summarized in Table 2.

Table 2. Results of simulation with one covariate (7000 replications)

2 dose levels	Correlation				
Group Size	0	0.3	0.5	0.7	0.9
20	0.044	0.044	0.044	0.046	0.050
40	0.048	0.046	0.043	0.049	0.049
120	0.043	0.046	0.044	0.043	0.049
3 dose levels	Correlation				
Group size	0	0.3	0.5	0.7	0.9
20	0.045	0.047	0.049	0.044	0.049
40	0.050	0.046	0.046	0.046	0.047
120	0.048	0.045	0.050	0.054	0.045

Since the entries in Table 2 do not differ significantly from 5% and tend to be on the conservative side of 5%, it is appropriate to use the proposed approximation for Williams' test also when a general linear model is being fitted.

REFERENCES

Brown MB and W Guo (1996). The choice of a test in a dose-response clinical trial: the effect of power. *Proceedings of the Biometrics Section, American Statistical Association,* pp 160-165.

Williams DA (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics* 27, 103-117.

Williams DA (1972). The comparison of several dose levels with a zero dose control. *Biometrics* 28, 519-531.

FRENCH RÉSUMÉ

Nous descrivons une approximation pour la statistique de Williams.