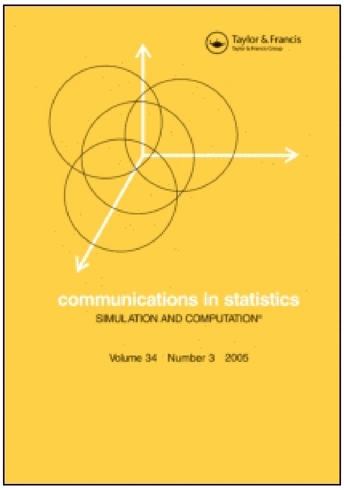

Communications in Statistics - Simulation and Computation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597237

ONE-STAGE AND TWO-STAGE STATISTICAL INFERENCE UNDER HETEROSCEDASTICITY

Shun-Yi Chen [a]
[a] Department of Mathematics, Tamkang University, Tamsui, Taiwan R.O.C.

Online Publication Date: 30 November 2001

COMMUN. STATIST.—SIMULA., 30(4), 991–1009 (2001)

ONE-STAGE AND TWO-STAGE STATISTICAL INFERENCE UNDER HETEROSCEDASTICITY

Shun-Yi Chen

Department of Mathematics, Tamkang University,
EA207, 151 Ying-Chuan Rd.,
Tamsui, Taiwan 251, R.O.C.
E-mail: sychen@math.tku.edu.tw

ABSTRACT

This paper presents general one-stage and two-stage sampling procedures for testing the hypotheses of the equality of means when the variances are unknown and unequal. The methods use a new range test and an ANOVA test based on independent t variables for an one-way layout. The one-stage procedure is an exact statistical procedure which provides a feasible solution to the two-stage procedure when the required two-stage sample sizes are not met due to time limitation, budget shortage, or other factors that result in early termination of the experiment. An extension to the two-way layout is also considered and the percentage points of the range test for testing the associated hypotheses are tabulated.

Key Words: Unknown and unequal variances; ANOVA; Range of independent t variables

1. INTRODUCTION

Consider I (≥ 2) independent populations $\pi_1, \ldots, \pi_I$ such that observations taken from population π_i are normally distributed with mean μ_i and variance σ_i^2 ($1 \leq i \leq I$). It is assumed that these variances are not known and possibly not equal, and that no prior knowledge about μ_i and σ_i^2 is available. This paper presents a new range test and an ANOVA test for testing the equality of means using a general one-stage and a two-stage sampling procedure.

The procedures for testing the equality of means in the conventional analysis of variance are based on the assumptions of normality, independence, and equality of the error variances. Studies have shown that the distribution of Tukey's range test or Fisher's F test depends heavily on the assumption of equal error variances, especially if the sample sizes are not equal (1). (2) pointed out that in practice not only the assumption of equal error variances is often unjustified, but there is also no exact theory to handle the case of heteroscedasticity. Historically transformations of the data, for example logarithm or arcsine transformations, are only approximate in terms of equal variances, normality, and model specification. These approximations, inherent in the transformation technique, lead to tests in the ANOVA setting which have only approximate level and power. It was also noted by (3) that if the errors associated with the original observations are normal, the errors associated with the transformed observations will not be normal. Furthermore, the transformed data may lose its practical meaning, and the acceptance of the equality of means using transformed data does not automatically imply the acceptance of the equality of means in the original scale. When the variances are unknown and unequal, (4) developed an exact analysis of variance for the means of I independent normal populations by using a two-stage procedure whose power is independent of the unknown variances. The number of samples that are required at the second stage can be large and may make the procedure impracticable. When the required sample size at the second stage is too large or when there is only one single sample available from each treatment, a modification of the two-stage procedure is needed. (5) and (6) developed a one-stage method for interval estimation and multiple comparisons, respectively. (3) used a one-stage sampling procedure to test the null hypotheses in ANOVA models under heteroscedasticity.

The one-stage procedure also provides an exact distribution for its test statistic under the null hypothesis, and it is both a data analysis-oriented procedure which works for any given set of data on hand and a design-oriented procedure to remedy the two-stage procedure. In Section 2 we present a general one-stage sampling procedure for testing the equality of means in the analysis of variance setting using a new range test and an ANOVA test. Section 3 presents comparisons of the new range test with the one-stage ANOVA test and with Tukey's studentized range test by simulation study. Simulation studies indicate that the new one-stage range test and the ANOVA test are equivalent. Simulation results also reveal that the one-stage range test is better than Tukey's studentized range test in terms of level of significance under unequal variances. In other words, Tukey's studentized range test has a level larger than the nominal one (60% to 200% larger), and hence it leads less likely to the rejection of the false null hypothesis. In Section 4, a two-stage procedure is briefly introduced, and a range test and an ANOVA test statistic are considered. It can be seen in Section 4.3 that the general one-stage procedure provides a practicable remedy to the two-stage procedure when its required sample sizes cannot be met. Therefore, the linkage between the two-stage and the one-stage procedures makes the former more applicable in design of experiments. Section 5 illustrates the use of the new range test with a numerical example. Extensions of the one-stage range test and the ANOVA test to the two-way model and the tables of the percentage points of the range test are given in Section 6.

2. THE GENERAL ONE-STAGE SAMPLING PROCEDURE

We now describe the general one-stage sampling procedure ($\boldsymbol{P}_1$) in the context of the one-way layout.

$\boldsymbol{P}_1$: Let X_{ij} $(j = 1, \dots, n_i)$ be an independent random sample of size n_i (≥ 3) from the normal population π_i $(i = 1, \dots, I)$ with unknown mean μ_i and unknown variance σ_i^2. For each population, employ the first (or any randomly chosen) n_0 $(2 \leq n_0 < n_i)$ observations to calculate the usual sample variance, denoted by S_i^2. Define weights U_i and V_i for the observations in the ith sample as

$$U_i = \frac{1}{n_i} + \frac{1}{n_i}\sqrt{\frac{n_i - n_0}{n_0}\left(n_i z^*/S_i^2 - 1\right)}$$
$$V_i = \frac{1}{n_i} - \frac{1}{n_i}\sqrt{\frac{n_0}{n_i - n_0}\left(n_i z^*/S_i^2 - 1\right)} \tag{2.1}$$

where z^* is the maximum of $\{S_1^2/n_1, \dots, S_I^2/n_I\}$. Note that U_i and V_i satisfy the following conditions

$$n_0 U_i + (n_i - n_0)V_i = 1, \quad [n_0 U_i^2 + (n_i - n_0)V_i^2]S_i^2 = z^*. \tag{2.2}$$

Calculate the final weighted sample mean using all observations by

$$\tilde{X}_{i.} = \sum_{j=1}^{n_0} U_i X_{ij} + \sum_{j=n_0+1}^{n_i} V_i X_{ij}, \tag{2.3}$$

which is a linear combination of the first set of observations with weights U_i and the second set of observations with weights V_i.

It was shown by (3) that the random variables

$$t_i = \frac{\tilde{X}_{i.} - \mu_i}{\sqrt{z^*}}, \quad i = 1, \ldots, I, \tag{2.4}$$

are distributed as independent Student's t with $\nu = n_0 - 1$ degrees of freedom.

From the above sampling procedure, (5) pointed out that the one-stage sample of size n_i is divided up into two portions, the first consisting of n_0 observations and the second consisting of the remaining $n_i - n_0$ observations. The final weighted sample mean $\tilde{X}_{i.}$ (with weights U_i associated with the first portion of observations and V_i with the second portion) depends on all S_i^2. The coefficient U_i is at least $1/n_i$ and V_i is at most $1/n_i$. Furthermore, the weight V_i will be negative if $z^* > S_i^2/n_0$.

When dealing with a situation where the data have already been collected, (5) suggested taking n_0 to be $n_i - 1$, because this choice pushes the weights U_i and V_i to be as close to $1/n_i$ as possible. Another reason of choosing $n_0 = n_i - 1$ is that it is optimal in the sense that the Student's t distribution has the smallest critical value (among $n_0 \leq n_i - 1$) for a fixed significance level. In situations where the one-stage procedure is used because the two-stage procedure cannot be completed due to uncontrollable factors such as budget shortage or early termination (see Section 4), one may choose n_0 to be between 10 and 15 as the two-stage procedure does at the initial stage (4). This choice of n_0 tends to stabilize the expected sample sizes for the two-stage procedure and tends to avoid the potentially large second stage sample sizes.

The model we are considering for the one-way layout is given by

$$X_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \ldots, I, \; j = 1, \ldots, n_i$$

where $\sum_{i=1}^{I} \alpha_i = 0$, and the e_{ij}'s are independent random errors with e_{ij} being distributed as $N(0, \sigma_i^2)$. We will use μ_i to denote $\mu + \alpha_i$. The goal is to test the null hypothesis that the population means (μ_i's) are all equal,

$$H_0 : \mu_1 = \cdots = \mu_I, \tag{2.5}$$

or equivalently $H_0 : \alpha_1 = \cdots = \alpha_I = 0$, against the alternative H_a : Not all μ_i's are equal. We consider two test statistics in the following subsections.

2.1. The One-Stage Range Test

Assume that for $i = 1, \ldots, I$, the one-stage sampling procedure has been conducted and that the final weighted sample means $\tilde{X}_{i.}$ have been computed as in (2.3). A new range test statistic for testing H_0 is proposed as follows:

$$T_1 = \frac{\tilde{X}_{\max} - \tilde{X}_{\min}}{\sqrt{z^*}} \tag{2.6}$$

where $\tilde{X}_{\max}(\tilde{X}_{\min})$ is the maximum (minimum) of $\tilde{X}_{1.}, \ldots, \tilde{X}_{I.}$. Under the null hypothesis H_0 in (2.5), it follows that T_1 is distributed as

$$R = \max_{1 \le i, j \le I} |t_i - t_j|$$

which is the range of I independent Student's t variates each with $\nu = n_0 - 1$ degrees of freedom. Thus, we reject the null hypothesis H_0 of equal population means if and only if $T_1 > R_{\alpha, I, \nu}$, where $R_{\alpha, I, \nu}$ is the upper α percentage point of the null distribution of T_1 which is the same as that of the random variable R. The percentage points for the range of I independent Student's t variates with ν degrees of freedom are given by (7) and (8).

2.2. The One-Stage ANOVA Test

For the one-stage sampling procedure ($\boldsymbol{P}_1$), (3) proposed the one-stage ANOVA test statistic given by

$$\tilde{F}^1 = \sum_{i=1}^{I} \frac{\left(\tilde{X}_{i.} - \tilde{X}_{..}\right)^2}{z^*} \tag{2.7}$$

where $\tilde{X}_{..} = \sum_{i=1}^{I} \tilde{X}_{i.}/I$. The null hypothesis of equal population means is rejected if and only if $\tilde{F}^1 > \tilde{F}_{\alpha, I, \nu}$, where $\tilde{F}_{\alpha, I, \nu}$ is the upper α percentage point of the null distribution of $\tilde{F}^1$ which can be found from (3), (9) and (10).

3. COMPARISON BETWEEN T_1 AND $\tilde{F}^1$

We compare the one-stage range test statistic T_1 in (2.6) with the one-stage ANOVA statistic $\tilde{F}^1$ in (2.7) through a simulation study. For the one-way layout with four independent populations ($I = 4$) and various sample sizes, powers of the test statistics T_1 and $\tilde{F}^1$ are reported in Table 1 by using Monte Carlo simulation. For each simulation run, under various configurations of variances and sample sizes, random variates of the standard normal distribution were generated using the random number generator RANNOR

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 1. Power Comparison of T_1 and $\tilde{F}^1$

n_i	n_0	T_1 Nominal Size 10%	5%	1%	$\tilde{F}^1$ Nominal Size 10%	5%	1%
(a) Variances: $1, 1, 1, 1$							
$(6, 6, 6, 6)$	5	0.245	0.130	0.021	0.244	0.129	0.020
$(6, 6, 8, 10)$	5	0.282	0.148	0.028	0.282	0.147	0.024
$(11, 11, 11, 11)$	10	0.606	0.453	0.189	0.613	0.469	0.194
$(6, 10, 16, 20)$	5	0.353	0.206	0.039	0.350	0.201	0.035
$(21, 21, 21, 21)$	20	0.938	0.887	0.705	0.937	0.884	0.691
(b) Variances: $1, 4, 4, 9$							
$(6, 6, 6, 6)$	5	0.122	0.062	0.012	0.121	0.062	0.011
$(6, 6, 8, 10)$	5	0.132	0.066	0.013	0.130	0.066	0.013
$(11, 11, 11, 11)$	10	0.187	0.109	0.024	0.191	0.110	0.026
$(6, 10, 16, 20)$	5	0.168	0.083	0.015	0.168	0.082	0.014
$(21, 21, 21, 21)$	20	0.307	0.196	0.060	0.311	0.198	0.062

in SAS 6.12 (11). The values of the test statistics T_1 and $\tilde{F}^1$ were calculated, and compared with the percentage points by using the tables of (8) and (10). Then the level and power of each statistic were found after 10,000 simulation runs. The levels of these two tests at $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0, 0, 0, 0)$, where $\alpha_i = \mu_i - \mu$, are very close to the nominal size for all cases. In Table 1 we reported the power obtained at $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (\sqrt{1/2}, -\sqrt{1/2}, 0, 0)$. For example, in the case where the variances are $(1, 4, 4, 9)$ and sample sizes are $(6, 10, 16, 20)$, we choose $n_0 = 5$; at the nominal 5% level, the calculated levels of T_1 and $\tilde{F}^1$ are both 0.05, and their powers are 0.083 and 0.082, respectively. Both for equal and unequal variances, the results of Table 1 indicate that the new one-stage range test performs as equally good as the one-stage ANOVA test. We have also compared these two tests when the number of populations (I) is 6 and 10, for various combinations of sample sizes and population variances. This leads to the same conclusion.

We also compare the one-stage range test T_1 with Tukey's studentized range test T_U (12), which is based on the equal variances assumption. Assuming the design to be balanced, i.e., $n_1 = \cdots = n_I = n$, $T_U = \sqrt{n}(\bar{X}_{\max} - \bar{X}_{\min})/S_p$, where $\bar{X}_{\max}$ ($\bar{X}_{\min}$) is the maximum (minimum) of $\bar{X}_{1\cdot}, \ldots, \bar{X}_{I\cdot}$, and $\bar{X}_{i\cdot} = \sum_{j=1}^{n} X_{ij}/n, i = 1, \ldots, I$; S_p^2 is the usual pooled sample variance estimate of the common variance σ^2. With four independent populations ($I = 4$) and various sample sizes in the one-way layout, the true levels of these test statistics T_U and T_1 (based on a Monte Carlo

Table 2. Comparison of T_1 and T_U

		T_1 Nominal Size		T_U Nominal Size	
n	n_0	5%	1%	5%	1%
(a) variances: $1, 3, 9, 27$					
6	5	0.05	0.01	0.080	0.032
11	10	0.05	0.01	0.080	0.032
21	20	0.05	0.01	0.078	0.031
41	40	0.05	0.01	0.076	0.028
61	60	0.05	0.01	0.082	0.031
(b) variances: $1, 4, 4, 9$					
6	5	0.05	0.01	0.070	0.018
11	10	0.05	0.01	0.065	0.020
21	20	0.05	0.01	0.065	0.020
41	40	0.05	0.01	0.064	0.022
61	60	0.05	0.01	0.073	0.020

simulation and using $n_0 = n - 1$) are reported in Table 2. For each simulation run, the values of these test statistics T_1 and T_U were calculated, and compared with their respective critical values. Then the level of each test was found after 10,000 simulation runs. The results of Table 2 indicate that the one-stage range test matches the nominal levels for the two variance combinations, while Tukey's studentized range test yields much higher p-values (60% to 200% higher) than the specified nominal levels. We have also compared the levels of two tests for $I = 6$ and 10 populations, and various combinations of sample sizes and variances, and found similar results as reported in Table 2. In other words, the one-stage range test performs better than Tukey's studentized range test when the variances are unequal. Tukey's studentized range test is less likely to reject the false null hypothesis because it yields a larger p-value than the nominal size. It should be noted that the variances in practice are unknown, the power based on Tukey's range test is not calculable unless they are specified, whereas the one-stage procedure can always provide an estimated power even if the variances are not known.

4. TWO-STAGE PROCEDURE

The two-stage sampling procedure ($\boldsymbol{P}_2$) proposed by (4) for a test of equality of means is briefly described as follows:

$\boldsymbol{P}_2$: Choose a number $z > 0$ (z is determined by the power of the test and will be discussed later), and take an initial sample of size n_0 (at least 2, but 10 or more will give better results) from each of the I populations. For the ith population let S_i^2 be the usual unbiased estimate of σ_i^2 based on the first n_0 observations, and define

$$N_i = \max\left\{n_0 + 1, \left[\frac{S_i^2}{z}\right] + 1\right\} \tag{4.1}$$

where $[x]$ denotes the greatest integer less than or equal to x. Then, take $N_i - n_0$ additional observations from the ith population so that we have a total of N_i observations from this population. We denote these by $X_{i1}, \ldots, X_{in_0}, \ldots, X_{iN_i}$. For each i, define

$$b_i = \frac{1}{N_i}\left[1 + \sqrt{\frac{n_0(N_i z - S_i^2)}{(N_i - n_0)S_i^2}}\right],$$

$$a_i = \frac{1 - (N_i - n_0)b_i}{n_0},$$

$$a_{i1} = \cdots = a_{in_0} = a_i,$$

$$a_{i,n_0+1} = \cdots = a_{iN_i} = b_i.$$

Now compute the weighted mean

$$\tilde{X}_{i.} = a_i \sum_{j=1}^{n_0} X_{ij} + b_i \sum_{j=n_0+1}^{N_i} X_{ij}$$

which is a linear combination of the first-stage data $(X_{i1}, \ldots, X_{in_0})$ and the second-stage data $(X_{i,n_0+1}, \ldots, X_{iN_i})$. It was shown that the random variables $t_i = (\tilde{X}_{i.} - \mu_i)/\sqrt{z}, i = 1, \ldots, I,$ are independently distributed following the t-distribution with $n_0 - 1$ d.f. (13). A reasonable choice of n_0 is between 10 and 15 for the initial sample size as suggested by (4). Such a choice can avoid a undesirable huge second stage sample size. Based on the weighted means a range test and an ANOVA test are proposed in the following.

4.1. Two-Stage Range Test

(14) proposed a two-stage range test to test the null hypothesis $H_0\colon \mu_1 = \cdots = \mu_I$ against the alternative $H_a\colon \mu_{\max} - \mu_{\min} \geq \delta^* > 0,$ where

$\mu_{\max}$ ($\mu_{\min}$) is the maximum (minimum) of $\mu_1, \ldots, \mu_I$. The range test statistic is given by

$$T_2 = \frac{\tilde{X}_{\max} - \tilde{X}_{\min}}{\sqrt{z}}$$

where $\tilde{X}_{\max}$ ($\tilde{X}_{\min}$) is the maximum (minimum) of $\tilde{X}_{1.}, \ldots, \tilde{X}_{I.}$. The null hypothesis H_0 is rejected at level α if and only if $T_2 > R_{\alpha, I, \nu}$, where $R_{\alpha, I, \nu}$ is the upper α percentage point of the range of I independent Student's t variates each with $\nu = n_0 - 1$ degrees of freedom, which can be found from Table 1 of (8).

The two-stage procedure is a design-oriented procedure which yields a test with controllable level and power. From (4.1) we can see that the final sample size depends on the first-stage data and z, where z depends on the predetermined alternative H_a and power. The power- and δ^*-related z values can be obtained from tables given by (14).

4.2. Two-Stage ANOVA Test

(4) proposed the ANOVA test statistic for testing $H_0 : \mu_1 = \cdots = \mu_I$ against $H_a : \sum(\mu_i - \bar{\mu}_.)^2 \geq \delta > 0$:

$$\tilde{F}^2 = \sum_{i=1}^{I} \frac{(\tilde{X}_{i.} - \tilde{X}_{..})^2}{z} \tag{4.2}$$

where $\bar{\mu}_. = \sum \mu_i / I$ and $\tilde{X}_{..}$ is the arithmetic mean of $\tilde{X}_{1.}, \ldots, \tilde{X}_{I.}$. The null hypothesis H_0 is rejected at level α if and only if $\tilde{F}^2 > \tilde{F}_{\alpha, I, \nu}$, where $\nu = n_0 - 1$, and the upper α percentage point $\tilde{F}_{\alpha, I, \nu}$ was discussed in Section 2.2. Tables of power- and δ-related z values can be found from (4).

Table 3 presents the comparison of z values for T_2 and $\tilde{F}^2$ when $I = 3, 4, 6, 10, 20$; $\alpha = 10\%, 5\%, 1\%$; $n_0 = 10, 15$; and power $(P^*) = .40, .90, .95$, under the same least favorable configuration of means subject to H_a. The two-stage range test T_2 and ANOVA test $\tilde{F}^2$ are equivalent because they produce almost the same z values (obtained by (14) and (4)), and hence the same sample sizes at given level and power. Since the two-stage range test is simpler in form and it is easy to interpret the alternative hypothesis H_a: $\mu_{\max} - \mu_{\min} \geq \delta^*$, we recommend the use of the range test in practice.

Table 3. Comparison of z-value for T_2 and $\tilde{F}^2$

			T_2 Level			$\tilde{F}^2$ Level		
I	P^*	n_0	10%	5%	1%	10%	5%	1%
3	.40	10	.283	.182	.092	.282	.181	.088
		15	.318	.208	.111	.318	.211	.111
	.90	10	.076	.061	.040	.075	.060	.039
		15	.083	.067	.046	.083	.067	.046
	.95	10	.061	.049	.034	.060	.049	.033
		15	.066	.055	.039	.066	.055	.039
4	.40	10	.227	.149	.078	.238	.154	.078
		15	.259	.173	.095	.262	.175	.093
	.90	10	.065	.053	.036	.067	.053	.035
		15	.073	.060	.042	.073	.059	.041
	.95	10	.054	.044	.031	.054	.044	.030
		15	.059	.049	.036	.059	.049	.035
6	.40	10	.172	.115	.063	.175	.116	.060
		15	.200	.137	.079	.202	.137	.076
	.90	10	.056	.045	.031	.055	.044	.029
		15	.062	.051	.037	.060	.050	.035
	.95	10	.046	.038	.027	.044	.037	.025
		15	.051	.043	.032	.050	.041	.030
10	.40	10	.127	.088	.050	.129	.087	.047
		15	.150	.106	.064	.154	.106	.061
	.90	10	.045	.037	.026	.043	.035	.024
		15	.052	.043	.032	.049	.041	.029
	.95	10	.038	.032	.023	.036	.030	.021
		15	.043	.037	.028	.041	.034	.025
20	.40	10	.086	.063	.038	.089	.061	.034
		15	.107	.079	.050	.106	.075	.044
	.90	10	.036	.030	.021	.032	.026	.018
		15	.042	.035	.027	.036	.031	.022
	.95	10	.030	.026	.019	.027	.023	.016
		15	.035	.030	.023	.031	.026	.020

4.3. Relation to One-Stage Procedure

With the two-stage tests T_2 and $\tilde{F}^2$ discussed in Sections 4.1 and 4.2 we can control both the level and the power despite unequal and unknown variances. The procedure is a useful design-oriented statistical method in an experiment. However, in situations where the experiment is terminated early due to budget restrictions, time limitations, or other uncontrollable factors, the required sample size N_i in (4.1) in the two-stage procedure cannot be completed. In this situation one can employ the one-stage procedure for the available data: use the first n_0 observations for the initial estimation and then use the total available data on hand to calculate the coefficients U_i and V_i in (2.1). Thus, given the sample data, the minimum power can be determined by the calculated probability at the least favorable configuration of means subject to H_a as follows:

$$P\left(\tilde{X}_{\max} - \tilde{X}_{\min} > \sqrt{z^*}\, R_{\alpha, I, \nu} | H_a\right) \quad \text{for the range test,}$$

and

$$P\left(\sum_{i=1}^{I} \left(\tilde{X}_{i.} - \tilde{X}_{..}\right)^2 > z^*\, \tilde{F}_{\alpha, I, \nu} | H_a\right) \quad \text{for the ANOVA test} \tag{4.3}$$

where $z^* = \max_{1 \leq i \leq I}(S_i^2/n_i)$. The power so determined is a data-dependent power which can be larger than, equal to, or smaller than the originally specified one. This is elaborated as follows. For $S_i^2 > zn_0,\ i = 1, \ldots, I$:

Case 1. If $S_i^2/n_i = S_j^2/n_j$ for all $i,\ j,$ then the two-stage (T_2 and $\tilde{F}^2$) and one-stage procedures (T_1 and $\tilde{F}^1$) can guarantee the same required power subject to the alternative hypothesis H_a. The power can be calculated by using existing tables by (4) for the ANOVA test and (14) for the range test.

Case 2. If $z^* < z,$ then the one-stage procedure has a power larger than that of the two-stage procedure.

Case 3. If $\min_{1 \leq i \leq I}(S_i^2/n_i) > z,$ then the power of the one-stage procedure is smaller than that of the two-stage procedure.

Case 4. In all other situations, the one-stage procedure could be better than, worse than, or equal to the two-stage procedure depending on the actual samples and the true population variances.

5. A NUMERICAL EXAMPLE

The data used in this section are from an experiment reported in (4) for studying the bacterial killing ability of four solvents. There were four types of solvents which can affect the ability of the fungicide methyl-2-benzimidazole-carbamate to destroy the fungus Penicillium expansum. The fungicide was diluted in exactly the same manner in four different types of solvents and sprayed on the fungus. The percentage of fungus destroyed was measured and reported. Let μ_i denote the mean percentage of fungus destroyed by solvent i. The aim of the experiment was to test the hypothesis that the mean percentages of fungus destroyed are all equal, $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$. (4) tried to use a two-stage sampling procedure for testing the hypothesis. At the first stage they collected $n_0 = 15$ observations for each solvent. The variances based on the first 15 observations were shown to be significantly different at the 5% level of significance by (6) using Bartlett's χ^2 test for equality of variances. The desired significance level for testing H_0 was .10, with a power of .85 under the alternative hypothesis $H_a: \mu_1 = \sqrt{1/2}, \mu_2 = -\sqrt{1/2}, \mu_3 = 0, \mu_4 = 0$. (The range of the parameters is $\sqrt{2}$.) So at their second stage the remaining $N_i - 15$ observations were taken according to the two-stage sampling procedure ($\boldsymbol{P}_2$) either by T_2 or $\tilde{F}^2$. In order to meet these requirements, the z value is found to be 0.0847 using the two-stage T_2 test, and the final sample sizes required for each solvent were 25, 38, 70 and 16, respectively (see (14) for details). However, due to some uncontrollable factors the experiment was *terminated* early, and the samples that were actually obtained were of size $N_1 = 24, N_2 = 36, N_3 = 67$, and $N_4 = 16$ (see Table 4A). In this situation one can apply the one-stage procedure ($\boldsymbol{P}_1$) as described in Section 2. Use the $n_0 = 15$ first-stage observations from each solvent for initial estimation and all the data in the final computation. The intermediate statistics S_i^2, U_i, V_i, and z^* are given in Table 4B. The final weighted sample means according to formula (2.3) are $\tilde{X}_{1.} = 97.429$, $\tilde{X}_{2.} = 94.905$, $\tilde{X}_{3.} = 94.862$, and $\tilde{X}_{4.} = 97.234$, respectively. We find the range test statistic $T_1 = 8.651$, which exceeds the 10% critical value $R_{.10,4,14} = 3.54$ at $\nu = 14$ degrees of freedom (obtained from Table 1 of (8)). The one-stage ANOVA test statistic $\tilde{F}^1 = 68.280$ exceeds the 10% critical value of 7.44 at $\nu = 14$ (from Table 1 of (4)). So H_0 is rejected by both tests. The estimated power for the specified alternative using T_1 is .83 according to $z^* = .08808$ by Case 4 of Section 4.3 and tables of (14) by interpolation. This power is slightly smaller than the .85 specified for the two-stage procedure, but fewer observations are available.

6. THE TWO-WAY MODEL

It is natural to extend the one-stage procedure to the two-way model of analysis of variance which is defined by:

$$
\begin{aligned}
X_{ijk} &= \mu_{ij} + e_{ijk} \\
&= \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk},
\end{aligned}
\tag{6.1}
$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 4A. Bacterial Killing Ability Data Used in the One-Stage Procedure

Solvent 1	Solvent 2		Solvent 3		Solvent 4	
96.44	93.63	94.03	93.58	93.43	92.42	97.18
96.87	93.99	92.43	93.02	92.72	92.38	97.42
97.24	94.61	92.62	93.86	93.56	92.06	97.65
95.41	91.69	94.47	92.90	94.13	92.50	95.90
95.29	93.00	94.14	91.43	93.57	92.54	96.35
95.61	94.17	93.09	92.68	96.27	92.52	97.13
95.28	92.62		91.57	98.05	91.80	96.06
94.63	93.41		92.87	97.67		96.33
95.58	94.67		92.65	98.93		96.71
98.20	95.28		95.31	97.23		98.11
98.29	95.13		95.33	95.95		98.38
98.30	95.68		95.17	97.79		98.35
98.65	97.52		98.59	97.41		98.05
98.43	97.52		98.00	96.94		98.25
98.41	97.37		98.79	97.08		98.12
98.59	96.97		96.36	98.15		97.97
98.20	97.21		96.69	96.73		
98.37	97.44		96.89	97.55		
98.57	96.86		96.13	94.44		
98.42	97.26		97.65	93.61		
98.29	98.27		97.81	93.61		
98.51	97.57		97.71	94.20		
98.89	97.81		97.48	94.20		
98.66	98.20		97.96	93.34		
	93.92		94.30	93.33		
	93.86		93.29	93.51		
	92.57		94.21	93.91		
	93.32		92.90	94.05		
	92.15		93.02	93.76		
	92.09		93.43	93.76		

Table 4B. Intermediate Statistics

	Solvent 1	Solvent 2	Solvent 3	Solvent 4
S_i^2	2.10995	3.17085	5.88428	0.77969
U_i	0.04306	0.02778	0.01642	0.07700
V_i	0.03934	0.02778	0.01449	−0.15501
$\tilde{X}_i$	97.4290	94.9047	94.8616	97.2339

$z^* = 0.08808$, $T_1 = 8.651$, $\tilde{F}^1 = 68.280$.

where $i = 1, \ldots, I$, $j = 1, \ldots, J, k = 1, \ldots, n_{ij}$. The e_{ijk}'s are independent random errors with e_{ijk} being distributed as $N(0, \sigma_{ij}^2)$, where the σ_{ij}^2 are unknown and possibly unequal, and we assume that

$$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{i=1}^{I} \alpha\beta_{ij} = \sum_{j=1}^{J} \alpha\beta_{ij} = 0.$$

The null hypotheses we wish to test are

$$H^1 : \alpha_i = 0 \quad \text{for all } i,$$

$$H^2 : \beta_j = 0 \quad \text{for all } j, \tag{6.2}$$

and

$$H^3 : \alpha\beta_{ij} = 0 \quad \text{for all } i \text{ and } j.$$

There are $I * J$ possible treatment combinations in the model (6.1). By cell (i, j) we mean the combination of level i of the first factor and level j of the second factor. The sample size for cell (i, j) is n_{ij} (≥ 3). The one-stage sampling procedure ($\boldsymbol{P}_3$) for testing the hypotheses of (6.2) proceeds as follows.

$\boldsymbol{P}_3$: Initially we employ the first (or randomly chosen) n_0 ($2 \leq n_0 < n_{ij}$) observations within each cell and compute the usual unbiased sample variance,

$$S_{ij}^2 = \sum_{k=1}^{n_0} (X_{ijk} - \bar{X}_{ij.})^2/(n_0 - 1).$$

where $\bar{X}_{ij.}$ is the cell sample mean based on size of n_0. Then the weights of the observations in cell (i, j) are chosen as

$$U_{ij} = \frac{1}{n_{ij}} + \frac{1}{n_{ij}} \sqrt{\frac{n_{ij} - n_0}{n_0} (n_{ij} z^*/S_{ij}^2 - 1)}$$

$$V_{ij} = \frac{1}{n_{ij}} - \frac{1}{n_{ij}} \sqrt{\frac{n_0}{n_{ij} - n_0} (n_{ij} z^*/S_{ij}^2 - 1)} \tag{6.3}$$

where z^* is the maximum value of $\{S_{ij}^2/n_{ij}, i = 1, \ldots, I, \ j = 1, \ldots, J\}$. Let the final weighted sample mean for cell (i, j) be defined by

$$\tilde{X}_{ij.} = \sum_{k=1}^{n_0} U_{ij} X_{ijk} + \sum_{k=n_0+1}^{n_{ij}} V_{ij} X_{ijk}, \tag{6.4}$$

and compute the statistics

$$\tilde{X}_{i..} = \frac{1}{J}\sum_{j=1}^{J}\tilde{X}_{ij.}, \quad \tilde{X}_{.j.} = \frac{1}{I}\sum_{i=1}^{I}\tilde{X}_{ij.}, \quad \tilde{X}_{...} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\tilde{X}_{ij.}$$

It can be shown, by a similar derivation as in Section 2, that

$$t_{ij} = \frac{\tilde{X}_{ij.} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij})}{\sqrt{z^*}}, \tag{6.5}$$

for $i = 1, \ldots, I,\ j = 1, \ldots, J,$ are distributed as independent Student's t with $n_0 - 1$ degrees of freedom.

6.1. The One-Stage Range Test

The range test statistic we use for H^1 is

$$\begin{aligned} T_1 &= \max_{1 \le i, i' \le I}\left|\frac{\tilde{X}_{i..} - \tilde{X}_{i'..}}{\sqrt{z^*}}\right| \\ &= \max_{1 \le i, i' \le I}\left|\bar{t}_{i.} - \bar{t}_{i'.} + \frac{\alpha_i - \alpha_{i'}}{\sqrt{z^*}}\right|; \end{aligned} \tag{6.6}$$

for H^2 we use

$$\begin{aligned} T_2 &= \max_{1 \le j, j' \le J}\left|\frac{\tilde{X}_{.j.} - \tilde{X}_{.j'.}}{\sqrt{z^*}}\right| \\ &= \max_{1 \le j, j' \le J}\left|\bar{t}_{.j} - \bar{t}_{.j'} + \frac{\beta_j - \beta_{j'}}{\sqrt{z^*}}\right|; \end{aligned} \tag{6.7}$$

and for H^3 we define

$$\begin{aligned} T_3 &= \max_{\forall i,j}\left|\frac{\tilde{X}_{ij.} - \tilde{X}_{i..} - \tilde{X}_{.j.} + \tilde{X}_{...}}{\sqrt{z^*}}\right| \\ &= \max_{\forall i,j}\left|t_{ij} - \bar{t}_{i.} - \bar{t}_{.j} + \bar{t}_{..} + \frac{\alpha\beta_{ij}}{\sqrt{z^*}}\right|, \end{aligned} \tag{6.8}$$

where

$$\bar{t}_{i.} = \frac{1}{J}\sum_{j=1}^{J} t_{ij}, \quad \bar{t}_{.j} = \frac{1}{I}\sum_{i=1}^{I} t_{ij}, \quad \bar{t}_{..} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} t_{ij}.$$

Table 5. Upper α-Percentage Points of T_1 ($R_{\alpha,J,I,\nu}$)

J	I	ν	10%	5%	1%	J	I	ν	10%	5%	1%
2	2	2	3.53	4.91	10.5	4	2	2	2.67	3.65	7.37
2	3	2	4.72	6.42	13.3	4	3	2	3.60	4.82	9.66
2	4	2	5.65	7.68	15.8	4	4	2	4.24	5.62	11.0
2	6	2	7.15	9.64	19.6	4	6	2	5.28	7.04	14.2
2	2	4	2.23	2.80	4.16	4	2	4	1.61	1.96	2.80
2	3	4	2.87	3.42	4.89	4	3	4	2.05	2.43	3.33
2	4	4	3.28	3.88	5.39	4	4	4	2.31	2.68	3.59
2	6	4	3.84	4.47	6.11	4	6	4	2.67	3.06	4.02
2	2	6	2.00	2.43	3.41	4	2	6	1.42	1.71	2.32
2	3	6	2.53	2.96	3.90	4	3	6	1.77	2.05	2.64
2	4	6	2.86	3.28	4.24	4	4	6	2.01	2.29	2.90
2	6	6	3.28	3.72	4.75	4	6	6	2.29	2.57	3.18
2	2	8	1.88	2.27	3.10	4	2	8	1.34	1.61	2.15
2	3	8	2.38	2.76	3.59	4	3	8	1.68	1.93	2.46
2	4	8	2.68	3.05	3.85	4	4	8	1.88	2.14	2.66
2	6	8	3.06	3.44	4.24	4	6	8	2.14	2.38	2.87
2	2	10	1.84	2.21	3.00	4	2	10	1.30	1.56	2.08
2	3	10	2.30	2.65	3.39	4	3	10	1.62	1.86	2.36
2	4	10	2.59	2.94	3.66	4	4	10	1.82	2.06	2.53
2	6	10	2.94	3.28	4.01	4	6	10	2.07	2.30	2.75
3	2	2	3.02	4.16	8.67	6	2	2	2.26	3.05	6.13
3	3	2	4.06	5.49	11.2	6	3	2	3.02	4.02	7.96
3	4	2	4.79	6.46	13.1	6	4	2	3.55	4.70	9.20
3	6	2	5.98	7.93	15.6	6	6	2	4.41	5.81	11.6
3	2	4	1.85	2.27	3.27	6	2	4	1.32	1.61	2.25
3	3	4	2.36	2.80	3.91	6	3	4	1.67	1.95	2.62
3	4	4	2.68	3.14	4.30	6	4	4	1.88	2.17	2.85
3	6	4	3.11	3.59	4.79	6	6	4	2.17	2.46	3.17
3	2	6	1.63	1.99	2.72	6	2	6	1.16	1.39	1.87
3	3	6	2.06	2.40	3.11	6	3	6	1.45	1.67	2.13
3	4	6	2.32	2.66	3.40	6	4	6	1.63	1.85	2.32
3	6	6	2.66	3.00	3.72	6	6	6	1.86	2.08	2.53
3	2	8	1.54	1.85	2.50	6	2	8	1.09	1.31	1.74
3	3	8	1.94	2.24	2.85	6	3	8	1.38	1.57	1.99
3	4	8	2.18	2.48	3.08	6	4	8	1.54	1.74	2.14
3	6	8	2.48	2.77	3.39	6	6	8	1.74	1.94	2.33
3	2	10	1.49	1.79	2.41	6	2	10	1.06	1.27	1.68
3	3	10	1.88	2.16	2.74	6	3	10	1.32	1.52	1.92
3	4	10	2.11	2.39	2.96	6	4	10	1.49	1.67	2.05
3	6	10	2.40	2.66	3.21	6	6	10	1.69	1.87	2.24

[†]Exchange of I and J to obtain the percentage points of T_2, $R_{\alpha,I,J,\nu}$.

Table 6. Upper α-Percentage Points of T_3 $(R^*_{\alpha,J,I,\nu})$

J	I	ν	10%	5%	1%	J	I	ν	10%	5%	1%
2	2	2	1.74	2.46	5.27	3	4	2	5.56	7.85	17.5
2	3	2	2.78	3.86	8.29	3	6	2	7.44	10.6	23.5
2	4	2	3.51	4.95	10.9	4	4	2	7.09	10.1	22.3
2	6	2	4.61	6.51	14.3	4	6	2	9.55	13.6	30.1
3	3	2	4.40	6.16	13.6	6	6	2	12.9	18.4	41.2
2	2	4	1.12	1.40	2.08	3	4	4	2.76	3.28	4.83
2	3	4	1.66	2.00	2.89	3	6	4	3.29	3.93	5.77
2	4	4	1.95	2.33	3.32	4	4	4	3.22	3.82	5.65
2	6	4	2.32	2.77	4.04	4	6	4	3.89	4.63	6.88
3	3	4	2.37	2.81	4.04	6	6	4	4.70	5.62	8.43
2	2	6	0.99	1.21	1.68	3	4	6	2.33	2.66	3.44
2	3	6	1.45	1.70	2.31	3	6	6	2.69	3.05	3.98
2	4	6	1.68	1.95	2.58	4	4	6	2.66	3.04	3.99
2	6	6	1.97	2.27	2.98	4	6	6	3.09	3.51	4.63
3	3	6	2.04	2.34	3.04	6	6	6	3.60	4.09	5.42
2	2	8	0.95	1.14	1.55	3	4	8	2.16	2.43	3.04
2	3	8	1.37	1.60	2.06	3	6	8	2.48	2.76	3.47
2	4	8	1.57	1.81	2.32	4	4	8	2.45	2.74	3.44
2	6	8	1.82	2.08	2.64	4	6	8	2.81	3.13	3.92
3	3	8	1.92	2.17	2.73	6	6	8	3.22	3.59	4.46
2	2	10	0.92	1.11	1.49	3	4	10	2.08	2.33	2.85
2	3	10	1.33	1.53	1.96	3	6	10	2.37	2.62	3.21
2	4	10	1.52	1.74	2.20	4	4	10	2.34	2.61	3.21
2	6	10	1.75	1.98	2.47	4	6	10	2.67	2.96	3.63
3	3	10	1.85	2.08	2.57	6	6	10	3.02	3.33	4.10

If the cell sizes n_{ij} are all equal, i.e., $n_{ij} = n, i = 1, \dots, I,\ j = 1, \dots, J$, then $\sqrt{z^*}$ is replaced by $S_{\max}/\sqrt{n}$ in (6.6) – (6.8), where $S^2_{\max}$ is the maximum value of the S^2_{ij}'s, $i = 1, \dots, I,\ j = 1, \dots, J$.

Under the null hypotheses of (6.2), equations in (6.6)–(6.8) yield the test statistics

$$T_1 = \max_{1 \le i, i' \le I} \left| \bar{t}_{i.} - \bar{t}_{i'.} \right|, \tag{6.9}$$

$$T_2 = \max_{1 \le j, j' \le J} \left| \bar{t}_{.j} - \bar{t}_{.j'} \right|, \tag{6.10}$$

and

$$T_3 = \max_{\forall i, j} \left| t_{ij} - \bar{t}_{i.} - \bar{t}_{.j} + \bar{t}_{..} \right|. \tag{6.11}$$

We reject H^1, H^2, or H^3 if T_1, T_2, or T_3 exceeds the upper α percentage points $R_{\alpha,J,I,\nu}$, $R_{\alpha,I,J,\nu}$ or $R^*_{\alpha,I,J,\nu}$ of its null distributions, respectively, where $\nu = n_0 - 1$. The percentage points for T_1 and T_2 for degrees of freedom $\nu = 2, 4, 6, 8$ and 10; $\alpha = .10, .05$ and .01; and the combinations of I, $J = 2, 3, 4, 6$ can be obtained from Table 5. The table was obtained by Monte Carlo simulation. In each simulation run, Student's $t = Y/\sqrt{X/\nu}$ variates were calculated where Y is the random variate of the standard normal distribution generated from the random number generator RANNOR and X is the chi-squared random variate with ν degrees of freedom generated by the gamma random number generator RANGAM in SAS 6.12. Then the values of T_1 (or T_2) were computed for each run. After 10,000 simulation runs, all the T_1 (or T_2) values were ranked in ascending order. The 90th, 95th and 99th percentiles were used to estimate the upper α percentage points of 10%, 5% and 1%, respectively. This process was replicated 9 times. The average values of the 9 critical values are listed in Table 5. The simulation errors of the percentage points mostly occur at the second decimal when α is large and at the first decimal when α is small. This is due to the long tail of the t distribution when the df is small. Similarly, the critical values $R^*_{\alpha,I,J,\nu}$ of T_3 in (6.11) are given in Table 6 for $I, J = 2, 3, 4, 6$; $\nu = 2, 4, 6, 8, 10$; and $\alpha = .10$, .05 and .01.

6.2. The One-Stage ANOVA Test

The one-stage ANOVA test statistics for H^1, H^2 and H^3 were proposed by (3), which are equivalent to the range tests T_1, T_2 and T_3. Since the form of the one-stage range test is simpler to use, we recommend the use of the range test.

ACKNOWLEDGMENTS

The author wishes to thank Professor Hubert J. Chen at the University of Georgia for his careful reading of this manuscript, and an associate editor and referees for their valuable suggestions. Part of this research was supported by NSC Grant Number 87-2118-M-032-005 of the National Science Council, Taiwan, and Tamkang University Research Fund.

REFERENCES

1. Bishop, T. A. Some results on simultaneous inference for analysis of variance with unequal variances. Technometrics, **1979**, *21*, 337–340.
2. Bishop, T. A.; Dudewicz, E. J. Heteroscedastic ANOVA. Sankhya, **1981**, *43B*, 40–57.
3. Chen, S. Y.; Chen, H. J. Single-stage analysis of variance under heteroscedasticity. Communications in Statistics – Simulation and Computation **1998**, *27* (3), 641–666.
4. Bishop, T. A.; Dudewicz, E. J. Exact analysis of variance with unequal variances: test procedures and tables. Technometrics, **1978**, *20*, 419–430.
5. Chen, H. J.; Lam, K. Single-stage interval estimation of the largest normal mean under heteroscedasticity. Communications in Statistics – Theory and Methods, **1989**, *18* (10), 3703–3718.
6. Wen, M. J.; Chen, H. J. Single-stage multiple comparison procedures under heteroscedasticity. American Journal of Mathematical and Management Sciences, **1994**, *14*, 1–48.
7. Hochberg, Y. A modification of the T-method of multiple comparisons for a one-way layout with unequal variances. Journal of the American Statistical Association, **1976**, *71*, 200–203.
8. Wilcox, R. R. A table of percentage points of the range of independent *t*-variables. Technometrics, **1983**, *25*, 201–204.
9. Dudewicz, E. J.; Ramberg, J. S.; Chen, H. J. New tables for multiple comparisons with a control (unknown variances). Biometrische Zeitschrift, **1975**, *17*, 13–76.
10. Bishop, T. A.; Dudewicz, E. J.; Juritz, J.; Stephens, M. A. Percentage points of quadratic forms in student *t* variates. Biometrika, **1978**, *65*, 435–439.
11. SAS Institute Inc. *SAS Language: Reference*, *Version 6*, First Edition. Cary, NC, 1990.
12. Miller, R. G. *Simultaneous Statistical Inference, Second Edition*. Springer-Verlag, New York, 1980.
13. Dudewicz, E. J.; Dalal, S. R. Allocation of observations in ranking and selection with unequal variances. Sankhya, **1975**, *37B*, 28–78.
14. Chen, S. Y.; Chen, H. J. A range test for the equality of means when variances are unequal. American Journal of Mathematical and Management Sciences, **2000**, *20* (1), 145–170.

Received March 1999
Revised May 2001