MULTIPLE COMPARISONS AMONG MEANS

OLIVE JEAN DUNN

University of California, Los Angeles

Methods for constructing simultaneous confidence intervals for all possible linear contrasts among several means of normally distributed variables have been given by Scheffé and Tukey. In this paper the possibility is considered of picking in advance a number (say m) of linear contrasts among k means, and then estimating these m linear contrasts by confidence intervals based on a Student t statistic, in such a way that the overall confidence level for the m intervals is greater than or equal to a preassigned value. It is found that for some values of k, and for m not too large, intervals obtained in this way are shorter than those using the F distribution or the Studentized range. When this is so, the experimenter may be willing to select the linear combinations in advance which he wishes to estimate in order to have m shorter intervals instead of an infinite number of longer intervals.

1. INTRODUCTION

THERE has been considerable work done on the problem of finding simultaneous confidence intervals for a number of linear contrasts among several means for normally distributed variables. Scheffé [1] gives a method for constructing simultaneous confidence intervals for all possible linear contrasts among k means using the F distribution. Tukey's intervals for all possible linear contrasts among k means use the distribution of the Studentized range [2]. Each of these methods may be extended to give confidence intervals for all possible linear combinations of the k means, as opposed to linear contrasts only.

In this paper the possibility is considered of picking in advance a number (say m) of linear combinations among the k means, and then estimating these m linear combinations by confidence intervals based on a Student t statistic, so that the overall confidence level for the m intervals is greater than or equal to a preassigned value, $1 - \alpha$. It is possible that for some values of k, and for m not too large, intervals obtained in this way may be shorter in some sense than those using the F distribution or the Studentized range. If this is so, the experimenter may be willing to select the linear combinations in advance which he wishes to estimate in order to have m shorter intervals instead of an infinite number of longer intervals.

The purpose of this paper, then, is to suggest and evaluate a simple use of the Student t statistic for simultaneous confidence intervals for linear combinations among several means, and to see under what conditions these intervals apply. The study was actually made with linear contrasts in mind, since these are probably estimated more frequently than other linear combinations among means. The paper has been written, however, in terms of arbitrary linear combinations, in order to stress the fact that the method is not limited to contrasts.

The method given here is so simple and so general that I am sure it must have been used before this. I do not find it, however, so can only conclude that

perhaps its very simplicity has kept statisticians from realizing that it is a very good method in some situations. In any case, the users of statistics in the main seem unaware of it, so I feel that it is worth presenting.

2. THE METHOD

Let the k means be $\mu_1, \cdots, \mu_k$, and let the estimates for them be $\hat{\mu}_1, \cdots, \hat{\mu}_k$, which are normally distributed with means $\mu_1, \cdots, \mu_k$ and with variances $a_{ii}\sigma^2$, $i = 1, \cdots, k$; let the covariance between $\hat{\mu}_i$ and $\hat{\mu}_j$ be $a_{ij}\sigma^2$ for $i \neq j$. Here the a_{ii} and a_{ij} are assumed to be known, but σ^2 may be unknown. Let $\hat{\sigma}^2$ be an estimate of σ^2 which is statistically independent of $\hat{\mu}_1, \cdots, \hat{\mu}_k$ and such that $\nu\hat{\sigma}^2/\sigma^2$ follows a Chi-square distribution with ν degrees of freedom. (These conditions are exactly those used by Scheffé in obtaining his intervals [1]. The condition that the dispersion matrix of the $\hat{\mu}_i$ be known except for a factor σ^2 is necessary in order to construct t statistics which are free of nuisance parameters.)

Let the m linear combinations of the means which are to be estimated be:

$$\theta_s = c_{1s}\mu_1 + \cdots + c_{ks}\mu_k, \qquad s = 1, 2, \cdots, m. \tag{1}$$

A linear combination is, in particular, a linear contrast if

$$\sum_{i=1}^{k} c_{is} = 0.$$

The unbiased estimates for $\theta_1, \cdots, \theta_m$ are

$$\hat{\theta}_s = c_{1s}\hat{\mu}_1 + \cdots + c_{ks}\hat{\mu}_k, \qquad s = 1, 2, \cdots, m. \tag{2}$$

These are m normally distributed variables, and the variance of $\hat{\theta}_s$ is $b_s^2\sigma^2$, where

$$b_s^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}c_{is}c_{js}.$$

This reduces the problem to one I discussed earlier [3], of finding confidence intervals for the means of m normally distributed variables.

The variates $t_1, \cdots, t_m$, each of which follows a t distribution with ν degrees of freedom, are formed:

$$t_s = \frac{\hat{\theta}_s - \theta_s}{b_s\hat{\sigma}}, \qquad s = 1, \cdots, m. \tag{3}$$

The variates $t_1, t_2, \cdots t_m$ have some joint distribution function; using a Bonferroni inequality, one can obtain a lower limit to the probability that all the t_i's lie between $-c$ and $+c$ (where c is any positive constant) without knowing anything about this joint distribution except that all the marginals are Student t distributions.

Thus

$$P[-c < t_i < c \quad i = 1, 2, \cdots \, m] \geq 1 - 2m \int_c^\infty f^{(\nu)}(t)dt, \tag{4}$$

where $f^{(\nu)}(t)$ is the frequency function for a Student t variable with ν degrees of freedom.

If c is selected so that the right hand member of (4) equals $1 - \alpha$, then confidence intervals with level $1 - \alpha$ are obtained from

$$P\left[-c < \frac{\hat{\theta}_s - \theta_s}{b_s \hat{\sigma}} < c, \quad s = 1, \cdots, m\right] \geq 1 - \alpha. \tag{5}$$

They are

$$\hat{\theta}_s \pm c b_s \hat{\sigma}, \qquad s = 1, 2, \cdots, m. \tag{6}$$

Here the overall confidence level for the m linear combinations is $1 - \alpha$, where c is defined by

$$\int_c^\infty f^{(\nu)}(t) dt = \frac{\alpha}{2m}, \quad \text{and} \quad f^{(\nu)}(t)$$

is the frequency function of a Student t variable with ν degrees of freedom. Some or all of these linear combinations may of course be linear contrasts.

When $\hat{\mu}_1, \cdots, \hat{\mu}_k$ are the sample means $\bar{y}_1, \cdots, \bar{y}_k$, and when $\bar{y}_1, \cdots, \bar{y}_k$ are statistically independent, then $a_{ii} = l/n_i$, where n_i is the size of the sample for the $\bar{y}_i$'s; for $i \neq j$, $a_{ij} = 0$. The confidence intervals for $c_{1s}\mu_1 + \cdots + c_{ks}\mu_k$ become

$$(c_{1s}\bar{y}_1 + \cdots + c_{ks}\bar{y}_k) \pm c \sqrt{\sum_{i=1}^k c_{is}^2/n_i}\, \hat{\sigma}, \qquad s = 1, \cdots, m. \tag{7}$$

Table 1 gives values of c for $1 - \alpha = .95$ and for various values of ν and of m; Table 2 gives c for $1 - \alpha = .99$. These tables as well as the other tables appearing in this paper have been computed from *Biometrika Tables for Statisticians*, Pearson and Hartley [4].

3. COMPARISON WITH INTERVALS USING F DISTRIBUTION

Scheffé's intervals for any number of linear contrasts among k means are

$$\hat{\theta}_s \pm S b_s \hat{\sigma} \tag{8}$$

where $S^2 = (k-1)F_\alpha(k-1, \nu)$. Here $F_\alpha(k-1, \nu)$ is the $1 - \alpha$ point of the F distribution with $k-1$ and ν degrees of freedom, and the other symbols are defined as in Section 2.

When intervals for a number of linear combinations are desired (not restricting them to linear contrasts), the intervals are as given in (8), but with $S^2 = kF_\alpha(k, \nu)$.

Since the t intervals in (6) and F intervals in (8) are seen to be of exactly the same form, and require exactly the same assumptions, it is both easy and useful to compare them.

The main difference between them is that in the t-intervals, the set of linear combinations which are to be estimated must be planned in advance, whereas with Scheffé's intervals they may be selected after looking at the data, since Scheffé's method gives intervals for all possible llnear combinations of k means.

TABLE 1. VALUES OF c FOR $1-\alpha=.95$

$$\int_{-\infty}^{c} f^{(\nu)}(t)dt = 1 - \frac{.05}{2m}$$

m \ ν	5	7	10	12	15	20	24	30	40	60	120	∞
2	3.17	2.84	2.64	2.56	2.49	2.42	2.39	2.36	2.33	2.30	2.27	2.24
3	3.54	3.13	2.87	2.78	2.69	2.61	2.58	2.54	2.50	2.47	2.43	2.39
4	3.81	3.34	3.04	2.94	2.84	2.75	2.70	2.66	2.62	2.58	2.54	2.50
5	4.04	3.50	3.17	3.06	2.95	2.85	2.80	2.75	2.71	2.66	2.62	2.58
6	4.22	3.64	3.28	3.15	3.04	2.93	2.88	2.83	2.78	2.73	2.68	2.64
7	4.38	3.76	3.37	3.24	3.11	3.00	2.94	2.89	2.84	2.79	2.74	2.69
8	4.53	3.86	3.45	3.31	3.18	3.06	3.00	2.94	2.89	2.84	2.79	2.74
9	4.66	3.95	3.52	3.37	3.24	3.11	3.05	2.99	2.93	2.88	2.83	2.77
10	4.78	4.03	3.58	3.43	3.29	3.16	3.09	3.03	2.97	2.92	2.86	2.81
15	5.25	4.36	3.83	3.65	3.48	3.33	3.26	3.19	3.12	3.06	2.99	2.94
20	5.60	4.59	4.01	3.80	3.62	3.46	3.38	3.30	3.23	3.16	3.09	3.02
25	5.89	4.78	4.15	3.93	3.74	3.55	3.47	3.39	3.31	3.24	3.16	3.09
30	6.15	4.95	4.27	4.04	3.82	3.63	3.54	3.46	3.38	3.30	3.22	3.15
35	6.36	5.09	4.37	4.13	3.90	3.70	3.61	3.52	3.43	3.34	3.27	3.19
40	6.56	5.21	4.45	4.20	3.97	3.76	3.66	3.57	3.48	3.39	3.31	3.23
45	6.70	5.31	4.53	4.26	4.02	3.80	3.70	3.61	3.51	3.42	3.34	3.26
50	6.86	5.40	4.59	4.32	4.07	3.85	3.74	3.65	3.55	3.46	3.37	3.29
100	8.00	6.08	5.06	4.73	4.42	4.15	4.04	3.90	3.79	3.69	3.58	3.48
250	9.68	7.06	5.70	5.27	4.90	4.56	4.4*	4.2*	4.1*	3.97	3.83	3.72

* Obtained by graphical interpolation.

This is, of course, a considerable advantage for Scheffé's method. It is, however, possible in using the t-intervals to select as the intervals to be estimated a very large set of linear combinations which includes all those which might conceivably be of interest. Then, on looking at the data, one may decide on actually

TABLE 2. VALUES OF c FOR $1-\alpha=.99$

$$\int_{-\infty}^{c} f^{(\nu)}(t)dt = 1 - \frac{.01}{2m}$$

m \ ν	5	7	10	12	15	20	24	30	40	60	120	∞
2	4.78	4.03	3.58	3.43	3.29	3.16	3.09	3.03	2.97	2.92	2.86	2.81
3	5.25	4.36	3.83	3.65	3.48	3.33	3.26	3.19	3.12	3.06	2.99	2.94
4	5.60	4.59	4.01	3.80	3.62	3.46	3.38	3.30	3.23	3.16	3.09	3.02
5	5.89	4.78	4.15	3.93	3.74	3.55	3.47	3.39	3.31	3.24	3.16	3.09
6	6.15	4.95	4.27	4.04	3.82	3.63	3.54	3.46	3.38	3.30	3.22	3.15
7	6.36	5.09	4.37	4.13	3.90	3.70	3.61	3.52	3.43	3.34	3.27	3.19
8	6.56	5.21	4.45	4.20	3.97	3.76	3.66	3.57	3.48	3.39	3.31	3.23
9	6.70	5.31	4.53	4.26	4.02	3.80	3.70	3.61	3.51	3.42	3.34	3.26
10	6.86	5.40	4.59	4.32	4.07	3.85	3.74	3.65	3.55	3.46	3.37	3.29
15	7.51	5.79	4.86	4.56	4.29	4.03	3.91	3.80	3.70	3.59	3.50	3.40
20	8.00	6.08	5.06	4.73	4.42	4.15	4.04	3.90	3.79	3.69	3.58	3.48
25	8.37	6.30	5.20	4.86	4.53	4.25	4.1*	3.98	3.88	3.76	3.64	3.54
30	8.68	6.49	5.33	4.95	4.61	4.33	4.2*	4.13	3.93	3.81	3.69	3.59
35	8.95	6.67	5.44	5.04	4.71	4.39	4.3*	4.26	3.97	3.84	3.73	3.63
40	9.19	6.83	5.52	5.12	4.78	4.46	4.3*	4.1*	4.01	3.89	3.77	3.66
45	9.41	6.93	5.60	5.20	4.84	4.52	4.3*	4.2*	4.1*	3.93	3.80	3.69
50	9.68	7.06	5.70	5.27	4.90	4.56	4.4*	4.2*	4.1*	3.97	3.83	3.72
100	11.04	7.80	6.20	5.70	5.20	4.80	4.7*	4.4*	4.5*		4.00	3.89
250	13.26	8.83	6.9*	6.3*	5.8*	5.2*	5.0*	4.9*	4.8*			4.11

* Obtained by graphical interpolation.

computing intervals for only some of this set. Section 5 gives an example of this procedure.

A second difference between the methods is that the lengths of the t-intervals depend on m, the number of linear combinations, whereas with Scheffé's method the lengths depend on k, the number of means. It seems reasonable to suspect, then, that the t-intervals may be shorter for small m and large k, and

TABLE 3. VALUES OF c^2/S^2 FOR $1-\alpha=.95, .99$

k m	1−α=.95					1−α=.99				
	2	5	10	15	20	2	5	10	15	20
ν=7										
2	1.44	.49	.24	.16	.12	1.33	.52	.27	.18	.14
5	2.19	.74	.37	.25	.19	1.87	.73	.38	.26	.19
10	2.91	.99	.49	.33	.25	2.38	.93	.48	.33	.25
50	5.22	1.77	.88	.59	.44	4.00	1.56	.81	.55	.42
100	6.61	2.24	1.12	.75	.56	4.97	1.94	1.00	.68	.52
250	8.77	2.97	1.48	.99	.75	6.36	2.48	1.28	.88	.66
ν=20										
2	1.35	.51	.27	.19	.14	1.23	.56	.32	.23	.18
5	1.87	.71	.38	.26	.20	1.56	.71	.40	.29	.22
10	2.30	.87	.46	.32	.25	1.83	.84	.48	.33	.26
50	3.41	1.29	.69	.48	.37	2.57	1.17	.67	.48	.37
100	3.96	1.50	.80	.55	.43	2.84	1.30	.74	.53	.41
250	4.78	1.81	.97	.67	.51	3.34	1.53	.87	.62	.48
ν=∞										
2	1.31	.53	.30	.21	.17	1.19	.59	.36	.27	.22
5	1.73	.70	.39	.28	.22	1.44	.72	.44	.33	.26
10	2.06	.83	.47	.33	.26	1.63	.82	.50	.37	.30
50	2.82	1.14	.64	.46	.36	2.09	1.04	.64	.48	.38
100	3.15	1.28	.72	.51	.40	2.28	1.14	.70	.52	.42
250	3.60	1.46	.82	.58	.46	2.55	1.27	.78	.58	.47

that the F intervals may be shorter for large m and small k. This turns out to be true.

Perhaps the most appealing way of comparing the two methods, from the standpoint of the research worker, is on the basis of length. To do this, Table 3 gives values of c^2/S^2 for certain values of k and m, for $\nu=7$, 20 and ∞, and for $1-\alpha=.95$ and .99. Here S^2 is defined as $(k-1)F_\alpha(k-1, \nu)$, so that the table is applicable as it stands when linear contrasts are being estimated. The square root of c^2/S^2 is the ratio of the length of a t-interval compared to the length of the corresponding Scheffé interval. Thus for $1-\alpha=.95$, $\nu=\infty$, $m=50$ and $k=10$, one has $c^2/S^2=.64$. This means that if one wishes to estimate 50 linear contrasts among 10 means, that each of the 50 t-intervals is .8 times as long as

the corresponding interval using the F distribution. A second interpretation is that about 64 per cent as many observations are necessary using the t-intervals to obtain the same precision as using the F distribution.

If one is estimating linear combinations among means rather than simply linear contrasts, one enters the table with a k value increased by one. Thus for 50 linear combinations among 9 means, with $1-\alpha=.95$ and $\nu=\infty$, one has $c^2/S^2=.64$.

TABLE 4. VALUES OF m_S, THE MAXIMUM NUMBER OF LINEAR
CONTRASTS OF k MEANS FOR WHICH F INTERVALS ARE
LONGER THAN t-INTERVALS, $1-\alpha=.95$, $.99$*

k \ ν	$1-\alpha=.95$			$1-\alpha=.99$		
	7	20	∞	7	20	∞
2	0	0	0	0	0	0
3	2	3	3	3	3	4
4	5	7	9	6	10	13
5	10	17	24	12	23	36
6	16	33	55	20	46	99
7	26	63	129	32	104	241
8	37	110	281	49	19×10	59×10
9	53	189	614	71	38×10	14×10^2
10	71	30×10	126×10	100		316×10
11	95	56×10	27×10^2			696×10
12	123	89×10	56×10^2			149×10^2
13	158	14×10^2	108×10^2			310×10^2
14	190		223×10^2			694×10^2
15	26×10		426×10^2			150×10^3
16	30×10		872×10^2			312×10^3
17	38×10		182×10^3			66×10^4
18	45×10		329×10^3			12×10^5
19	54×10		635×10^3			26×10^5
20			132×10^4			6×10^6

* The last significant digit given in m_S cannot be expected to be exactly correct except where m_S is smaller than 100. For example, $m_S=26\times10^5$ indicates that at some point in the calculation a number with only two significant digits was used; in computing $n_S=36$, on the other hand, three significant digits were carried throughout.

It appears from Table 3 that for a fairly large number of means, the t-intervals are shorter for any m of reasonable size.

In Table 4 are listed values of m_S, the maximum number of linear combinations for which the Scheffé intervals are longer than the t intervals. If, for a given k, ν, and $1-\alpha$, one decides to estimate m linear contrasts among k means, one may examine m_S from Table 4. If $m\leq m_S$, the Student t intervals are shorter; if $m>m_S$ the Scheffé intervals are shorter. Table 4 gives m_S for $\nu=7$, 20, ∞, $1-\alpha=.95$ and $.99$, and $k=2, 3, \cdots, 20$. The table indicates that for k as large as 10, one may form a large number of contrasts using the t-intervals and still have intervals smaller than if the F distribution had been used.

If one wishes to estimate linear contrasts among 9 means with $\nu=20$, $1-\alpha=.95$, then entering Table 4 at $k=9$, one finds $m_S=189$, so that if the number

of contrasts to be estimated is less than or equal to 189, the Student t intervals should be used rather than Scheffé's intervals. If linear combination rather than just contrasts are being estimated, then one enters the table with 10 rather than 9 and finds $m_S = 300$.

Examination of Table 4 indicates that the situation becomes more favorable for the Student t method if, as all other variables except one are held constant:

1) k is increased; or
2) ν is increased; or
3) $1 - \alpha$ is increased.

4. COMPARISON WITH TUKEY'S INTERVALS

If $\hat{\mu}_1, \cdots, \hat{\mu}_k$ are unbiased, normally distributed estimators of $\mu_1, \cdots, \mu_k$ with Var $(\hat{\mu}_i) = a_{11}\sigma^2$ and Cov $(\hat{\mu}_i, \hat{\mu}_j) = 0$, then confidence intervals of level $1 - \alpha$ for all possible linear contrasts

$$\theta = c_1\mu_1 + \cdots + c_k\mu_k, \qquad \sum_{i=1}^{k} c_i = 0,$$

are given by

$$\hat{\theta} \pm \frac{1}{2} \sum_{i=1}^{k} |c_i| \sqrt{a_{11}}q\hat{\sigma}, \tag{9}$$

where q is the $1 - \alpha$ point of the Studentized range for a sample of size k, and $\hat{\sigma}^2$ is an independent estimate of σ^2 such that $\nu\hat{\sigma}^2/\sigma^2$ has a χ^2 distribution with ν degrees of freedom.

For $\hat{\mu}_i = \bar{y}_i$, based on a sample of size n, these become

$$\hat{\theta} \pm \sum_{i=1}^{k} |c_i| q\hat{\sigma}/\sqrt{n}. \tag{10}$$

It is apparent that, as formulated here, these intervals are more limited in application than the t intervals and Scheffé's intervals, since (1) they apply only to linear contrasts rather than to arbitrary linear combinations; (2) the variances are assumed to be equal; (3) the covariances are assumed to be all zero. Tukey [2], states without proof that they may be extended somewhat in all three directions.

Limitation (1) may be removed by introducing a $(k+1)$st mean whose estimate, $\hat{\mu}_{k+1}$, is always zero. Tukey shows that there is no appreciable error in simply using the same intervals for linear combinations as for linear contrasts, provided $k > 2$. To be ultra conservative, one may use the same intervals but enter the tables of the Studentized range with $k+1$ rather than with k.

In using Tukey's intervals for combinations as opposed to contrasts, it should be noted that formulas (9) and (10) must be altered by replacing $\frac{1}{2}\sum |c_i|$ by the larger of the sum of the positive $\{c_i\}$ and the negative of the sum of the negative $\{c_i\}$.

In an effort to remove limitation (2), Tukey considers the case where Var $(\hat{\mu}_i) = a_{ii}\sigma^2$, $i = 1, \cdots, k$, and Cov $(\hat{\mu}_i, \hat{\mu}_j) = 0$, $i \neq j$, with the a_{ii} known con-

stants. In other words, the covariances among the $\hat{\mu}_i$ are zero and the ratios among their variances are known.

By multiplying the $q\hat{\sigma}$ in (9) by a factor depending on the particular contrast and on the various a_{ii}, instead of by $\sqrt{a_{11}}$, he obtained intervals with an overall confidence level which he says is approximately $1 - \alpha$, adding that work is being done on studying this approximation.

Limitation (3) may also be removed when only contrasts are being considered. If Var $(\hat{\mu}_i) = a_{11}\sigma^2$, $i = 1, \cdots, k$, and Cov $(\hat{\mu}_i, \hat{\mu}_j) = a_{12}\sigma^2$, $i \neq j$, where a_{11} and a_{12} are known, then the confidence interval for

$$\theta = \sum_{i=1}^{k} c_i \mu_i$$

becomes

$$\hat{\theta} \pm \frac{1}{2} \sum_{i=1}^{k} |c_i| (a_{11} - a_{12})^{1/2} q\hat{\sigma}. \tag{11}$$

Thus the contrasts using the Studentized range have been partially extended, but not to the more general situation where the variances and covariances of the $\hat{\mu}_i$ are $a_{ii}\sigma^2$ and $a_{ij}\sigma^2$, with the a_{ii} and a_{ij} known. The extension to unequal variances seems to be somewhat arbitrary, and I do not know whether it has been put on any satisfactory basis.

Comparison of the lengths of the Studentized range intervals with the t-intervals is complicated by the fact that the lengths of the Tukey intervals depend on $\sum |c_i|$ whereas the lengths of the t intervals depend on

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} c_{is} c_{js}.$$

Scheffé [1] compares the squared length of an F-interval with the squared length of the same Studentized range interval, and then considers the maximum and minimum values of this ratio over all types of contrast. He points out that this squared ratio is a maximum for intervals of the type $\mu_i - \mu_{i'}$ and a minimum for intervals of the type

$$2(\mu_1 + \cdots + \mu_{k/2})/k - 2(\mu_{k/2+1} + \cdots + \mu_k)/k$$

for k even, or of the type

$$(\mu_1 + \cdots + \mu_{k/2-1}) \Big/ \left(\frac{k}{2} - 1\right) - (\mu_{k/2} + \cdots + \mu_k) \Big/ \left(\frac{k}{2} + 1\right)$$

for k odd.

In Table 5 are given values of c^2/q^2. From this table, for any particular contrast one may compute the squared ratio of lengths for the t-intervals and Studentized range intervals. For variances equal to $a_{11}\sigma^2$ and covariances zero, this squared ratio is

$$4\left[\sum_i c_{si}^2 \Big/ \left(\sum_i |c_{si}|\right)^2 \right] (c^2/q^2).$$

TABLE 5. VALUES OF c^2/q^2 FOR $1-\alpha=.95,\ .99$

m \ k	$1-\alpha=.95$					$1-\alpha=.99$				
	2	5	10	15	20	2	5	10	15	20
$\nu=7$										
2	.72	.32	.21	.18	.16	.66	.33	.23	.20	.17
5	1.10	.48	.32	.27	.24	.93	.46	.33	.27	.25
10	1.46	.63	.43	.36	.32	1.19	.59	.42	.35	.31
50	2.61	1.14	.77	.64	.57	2.00	1.00	.70	.59	.53
100	3.31	1.44	.97	.81	.72	2.48	1.24	.87	.73	.65
250	4.39	1.91	1.29	1.07	.95	3.18	1.59	1.11	.94	.84
$\nu=20$										
2	.67	.33	.23	.20	.18	.62	.36	.27	.23	.21
5	.93	.45	.32	.28	.25	.78	.45	.34	.30	.27
10	1.15	.56	.40	.34	.30	.92	.53	.40	.35	.32
50	1.70	.83	.59	.50	.45	1.29	.74	.56	.49	.45
100	1.98	.96	.69	.58	.53	1.43	.82	.62	.54	.50
250	2.39	1.16	.83	.71	.64	1.67	.97	.73	.64	.58
$\nu=\infty$										
2	.65	.34	.25	.22	.20	.60	.37	.30	.27	.25
5	.87	.45	.33	.29	.27	.72	.45	.36	.32	.30
10	1.03	.53	.40	.34	.31	.82	.51	.41	.36	.34
50	1.41	.73	.54	.47	.43	1.04	.65	.52	.47	.43
100	1.58	.81	.61	.53	.48	1.14	.72	.57	.51	.47
250	1.80	.93	.69	.60	.55	1.27	.80	.63	.57	.53

In Table 6 are given values of m_T, the maximum m such that *every* t interval is shorter than the corresponding Tukey interval, even for the least favorable (to the t interval) case, $\mu_i-\mu_{i'}$. A glance at Table 6 shows that m_T tends to be rather small. If one's primary interest is in intervals like $\mu_i-\mu_{i'}$, then he may use this table to decide which method to use. Otherwise a comparison must be made for each type of interval which is of interest.

Again, as in the comparison between Scheffé's intervals and the t intervals, the t intervals seem to become better, other things being equal, as

1) k becomes larger, or
2) ν becomes larger, or
3) $1-\alpha$ becomes larger.

In an analysis of variance situation with a single variable of classification, the number of means would tend to be small and primary interest might be in estimating the difference between means. Then Tukey's intervals are perhaps preferable. When there are two variables of classification, then one perhaps wishes to estimate row differences, column differences, and interactions rather than the differences between single means. Then the t intervals are more likely to be shorter.

TABLE 6. VALUES OF m_T, THE MAXIMUM NUMBER OF LINEAR CONTRASTS OF k MEANS FOR WHICH EVERY STUDENTIZED RANGE INTERVAL IS LONGER THAN THE CORRESPONDING t-INTERVAL, $1-\alpha=.95, .99$

k \ ν	$1-\alpha=.95$			$1-\alpha=.99$		
	7	20	∞	7	20	∞
2	0	0	0	0	0	0
3	2	2	2	2	2	2
4	3	4	4	4	4	5
5	5	6	7	6	7	8
6	7	9	11	8	11	13
7	9	12	15	10	14	17
8	11	16	20	12	18	24
9	13	20	25	14	23	30
10	14	24	31	17	28	38
11	16	28	39	19	33	46
12	18	33	47	21	39	54
13	20	37	53	24	43	63
14	22	43	61	26	50	74
15	24	48	71	29	56	84
16	26	53	82	31	64	95
17	28	60	92	33	71	112
18	30	66	103	35	80	122
19	32	71	111	37	92	139
20	34	77	125	39	104	157

5. AN EXAMPLE

As an example of an experiment in which one may wish to choose among these methods, consider the fixed-effect model for a two-way classification with a rows, b columns, and n observations in each cell.

Let x_{ijk} be the kth observation in the ith row and the jth column. Then $E(x_{ijk})=\mu_{ij}=\mu+a_i+b_j+I_{ij}$, and $\mathrm{Var}\,(\bar{x}_{ij.})=\sigma^2/n$ for $i=1,\,\cdots,a;\,j=1,\,\cdots,b;\,k=1,\,\cdots,n$, with

$$\sum_{i=1}^{a} a_i = 0, \qquad \sum_{j=1}^{b} b_j = 0, \qquad \sum_{i=1}^{a} I_{ij} = 0, \qquad j = 1,\,\cdots,b,$$

and

$$\sum_{j=1}^{b} I_{ij} = 0, \qquad i = 1,\,\cdots,a.$$

The pooled estimate of the variance, $\hat{\sigma}^2$, has $n(a-1)(b-1)$ degrees of freedom. Here $a_{ii}=1/n$, $a_{ij}=0$ for $i \neq j$, and if

$$\theta = \sum_{i,j} c_{ij}\mu_{ij}$$

is any linear combination of the ab means, the point estimate for it is

$$\hat{\theta} = \sum_{i,j} c_{ij}\bar{x}_{ij.},$$

and the confidence interval using the t-statistic is

$$\sum_{i,j} c_{ij} \bar{x}_{ij} \pm c \sqrt{\sum_{i,j} c^2_{ij}/n} \, \hat{\sigma}.$$

In the first column of Table 7 are listed various linear combinations which the experimenter may wish to estimate. For ease in comparison, they have been multiplied where necessary by a factor chosen so that the lengths of the Tukey intervals are all equal. In the second column are listed the usual point estimates. Those listed in rows 5 to 10 of the table are linear contrasts, whereas those in rows 1 to 4 are not. In the third column are given the number of each type of linear combination; the fourth column gives

$$\sum_{i,j} \overset{2}{c}_{ij}/n,$$

to be used in the confidence interval for that type of contrast.

The last two columns of Table 7 give

$$c \sqrt{\sum_{i,j} \overset{2}{c}_{ij}/n}$$

for $a = 3$, $b = 4$, $n = 3$ and for $a = 4$, $b = 5$, and $n = 4$. In these columns, c has been computed on the assumption that the experimenter wishes to estimate all the linear combinations listed, with an overall confidence level of $1 - \alpha = .95$. To compare the t-intervals with Tukey's intervals, the values in columns (5) and (6) must be compared with 2.99 and 2.64, respectively, since Tukey's intervals for all these linear combinations are $c_{ij} \bar{x}_{ij} \pm (q/\sqrt{n}) \hat{\sigma}$.

To compare the lengths of the t-intervals with Scheffé's intervals, one must compare the values of c, 3.97 and 4.31, with 5.11 and 5.92, the corresponding values of S.

It is possible that the experimenter may be interested in estimating not all the linear combinations. In Table 8 are shown values of c for various sets of linear combinations estimated: rows 1 to 10 inclusive (all the linear combinations listed); rows 1 to 9 inclusive (all the linear combinations listed except the differences between means); rows 5 to 10 inclusive (all the linear contrasts listed); and rows 5 to 9 inclusive (all the linear contrasts listed except the differences between means). Table 8 also gives the corresponding values of S.

In these particular examples one should probably pick the t-statistic over the Studentized range. The exception to this (considering only length of interval) would be if one is mainly interested in estimates of $\mu_{ij} - \mu_{i'j'}$; this does not seem likely. It should be emphasized, however, that if the research worker wants intervals as short as possible, each problem must be examined separately.

6. DISCUSSION OF THE STUDENT t METHOD

It is interesting to consider what may be the actual probability of coverage using the Student t method. In (4), we may let the left side (the actual probability of coverage) be denoted by P and the right side (the lower bound for this

TABLE 7. EXAMPLE OF TWO-WAY CLASSIFICATION USING t-INTERVALS
$(1-\alpha=.95)$

$\theta = \sum_{ij} c_{ij}\mu_{ij}$ (1)	$\hat{\theta} = \sum_{ij} c_{ij}\bar{x}_{ij\cdot}$ (2)	Number (3)	$\sum_{ij} \dfrac{c_{ij}^2}{n}$ (4)	$\begin{array}{c}a=3, b=4,\\ n=3\\ \nu=24\\ c\sqrt{\sum c_{ij}^2/n}\\ (m=84, c=3.97)\\ (5)\end{array}$	$\begin{array}{c}a=4, b=5,\\ n=4\\ \nu=60\\ c\sqrt{\sum c_{ij}^2/n}\\ (m=195, c=4.31)\\ (6)\end{array}$
1. μ	$\bar{x}_{\cdots}$	1	$4/nab$	.66	.48
2. $\mu + a_i$	$\bar{x}_{i\cdots}$	a	$1/nb$	1.14	.96
3. $\mu + b_j$	$\bar{x}_{\cdot j\cdot}$	b	$1/na$	1.32	1.08
4. μ_{ij}	$\bar{x}_{ij\cdot}$	ab	$1/n$	2.29	2.16
5. $\left(\dfrac{a}{a-1}\right)a_i$	$\left(\dfrac{a}{a-1}\right)(\bar{x}_{i\cdots} - \bar{x}_{\cdots})$	a	$\dfrac{a}{n(a-1)b}$	1.40	1.11
6. $\left(\dfrac{b}{b-1}\right)b_j$	$\left(\dfrac{b}{b-1}\right)(\bar{x}_{\cdot j\cdot} - \bar{x}_{\cdots})$	b	$\dfrac{b}{na(b-1)}$	1.53	1.21
7. $\dfrac{ab}{2(a-1)(b-1)}I_{ij}$	$\dfrac{cb}{2(a-1)(b-1)}(\bar{x}_{ij\cdot} - \bar{x}_{i\cdots} - \bar{x}_{\cdot j\cdot} + \bar{x}_{\cdots})$	ab	$\dfrac{ab}{4n(a-1)(b-1)}$	1.62	1.39
8. $a_i - a_{i'}$	$\bar{x}_{i\cdots} - \bar{x}_{i'\cdots}$	$a(a-1)/2$	$2/nb$	1.62	1.36
9. $b_j - b_{j'}$	$\bar{x}_{\cdot j\cdot} - \bar{x}_{\cdot j'\cdot}$	$b(b-1)/2$	$2/na$	1.87	1.52
10. $\mu_{ij} - \mu_{i'j'}$ $i \neq i', j \neq j'$	$\bar{x}_{ij\cdot} - \bar{x}_{i'j'\cdot}$	$ab(ab-1)/2$	$2/n$	3.24	3.05

TABLE 8. COMPARISON BETWEEN LENGTHS OF t-INTERVALS AND
F-INTERVALS IN TWO-WAY CLASSIFICATION EXAMPLE
$(1-\alpha=.95)$

Types Estimated	$a=3, b=4, n=3$			$s=4, b=5, n=4$		
	Number Estimated	c	S	Number Estimated	c	S
1 to 10, incl.	84	3.97	5.11	195	4.31	5.92
1 to 9, incl.	48	3.73	5.11	75	3.93	5.92
5 to 10, incl.	64	3.87	4.93	165	4.26	5.78
5 to 9, incl.	28	3.53	4.93	45	3.71	5.78

probability of coverage) be denoted by $\underline{P}$, and consider the difference between P and $\underline{P}$.

If all the correlations approach unity, then in the limit all the linear combinations become one and the same linear combination, and P attains its largest possible value. In this case,

$$P = P(-c < t_1 < c) = 1 - 2 \int_c^\infty f^{(\nu)}(t)dt.$$

For $\nu = \infty$, $\underline{P} = 1 - \alpha = .95$, and $m = 100$, $c = 3.48$, and $P = .9994$.

In [3] I conjectured that P attains its smallest possible value when all the correlations are zero. This was established, however, only for $m = 2$ and 3.

Extensive tables are not available at present to evaluate P when all the correlations are zero, though Pillai and Ramachandran [5] give 95 per cent points and 99 per cent points of the necessary distribution for m less than or equal to 8. For $\nu = \infty$, however, the normal tables may be used. For $\underline{P} = 1 - \alpha$, $P = [1 - (\alpha/m)]^m$, so that $\underline{P}$ is the first two terms in the binomial expansion of $[1 - (\alpha/m)]^m$. The difference between them is seen to be bounded by $[(m-1)/2m]\alpha^2$, so that for $\nu = \infty$ and α small, P is fairly close to the actual probability of coverage when all the correlations are zero. In particular, for $1 - \alpha = .95$ and $m = 100$, $P = (1 - .0005)^{100} = .9512$.

Thus for $\nu = \infty$ and α small, the inequality (4) gives results which are almost as good as any which are attainable when nothing is known about the correlations. At present I am attempting to construct tables which will give some idea of the situation for small values of ν and for correlations between 0 and 1.

REFERENCES

[1] Henry Scheffé, "A method of judging all contrasts in the analysis of variance," *Biometrika*, 40 (1953), 87–104.
[2] John M. Tukey, "The problem of multiple comparisons," mimeographed notes, Princeton University.
[3] Olive Jean Dunn, "Estimation of the means of dependent variables," *Annals of Mathematical Statistics*, 29 (1958), 1095–111.
[4] E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. 1, Cambridge, 1956.
[5] K. C. S. Pillai and K. V. Ramachandran, "Distribution of a Studentized Order Statistic," *Annals of Mathematical Statistics*, 25 (1954), 565–72.