SIMULTANEOUS TEST PROCEDURES—SOME THEORY OF MULTIPLE COMPARISONS

BY K. R. GABRIEL

University of North Carolina and Hebrew University, Jerusalem

1. Introduction and summary. When a hypothesis is tested by a significance test and is not rejected, it is generally agreed that all hypotheses implied by that hypothesis (its "components") must also be considered as non-rejected. However, when the hypothesis is rejected the question remains as to which components may also be rejected. Various writers have given attention to this question and have proposed a variety of multiple comparisons methods based either on tests of each one of the components or on simultaneous confidence bounds on parametric functions related to the various hypotheses.

An approach to such methods, apparently originally due to Tukey [27], is to test each component hypothesis by comparing its statistic with the \( \alpha \) level critical value of the statistic for the overall hypothesis. This is called a Simultaneous Test Procedure (STP for short) as all hypotheses may be tested simultaneously and without reference to one another. An STP involves no stepwise testing of the kind employed by some other methods of multiple comparisons for means, in which subsets are tested for equality only if they are contained in sets which have already been found significant. (See [3], [4], [10], [18]).

A general formalization of STP's is attempted in this paper. Section 2 introduces the requisite concepts of families of hypotheses and the implication relations between them, as well as the monotonicity relations between the related statistics. Section 3 defines STP's and shows conditions for coherence and consonance of their decisions, these properties being that hypothesis implication relations are preserved in the decisions of the STP. Section 4 discusses comparison of various STP's for the same hypotheses and shows the advantages of the union-intersection type of statistics and of reducing the family of hypotheses tested as much as possible. Section 5 translates all these results to simultaneous confidence statements after introducing the definitions necessary to allow such translation.

The analogy between simultaneous test and confidence methods is of special importance as it brings a wide spectrum of methods within this framework, most of which was originally formulated in confidence region terms. This covers the original work of Tukey [27] and Scheffé [25] and continues with that of Roy and his associates [21] and most recently Krishnaiah [12], [13]. A general discussion of this confidence approach has been given by Aitchison [1] since the first draft of the present paper. In view of the close analogies pointed out in Section 5, it is

Received 2 October 1963, revised 2 August 1967.

1 The research was supported in part by the National Institutes of Health (Institute of General Medical Sciences) Grant No. GM-12868-03, and in part by National Center for Health Statistics Grant No. NCHS-IS-1.
not surprising that Aitchison arrives at the requirement of a constant critical value for all his tests, exactly as in an STP. In fact his approach and the present one are complementary.

The formal theory is illustrated with ANOVA examples to clarify the concepts. No essentially new techniques are presented in this paper though this approach has been used elsewhere by the author to derive a number of practically useful procedures [6], [7], [8].

2. Families of hypotheses and statistics. Let \( Y \) be a (vector) random variable having density function \( f_\theta(Y) \) with respect to a \( \sigma \)-finite measure \( \mu \), where the parameter(s) are assumed to lie in a set \( \omega \). An overall hypothesis \( \omega_0 (\subset \omega) \) is being considered, as well as a family \( \Omega = \{ \omega_i | i \in I \} \) of hypotheses \( \omega_i \) implied by \( \omega_0 \), i.e., \( \omega_0 \subseteq \omega_i \subseteq \omega \) if \( i \in I \), \( I \) being an index set, not necessarily denumerable. (Note that \( \omega_i \) is being used to denote both a set in parameter space \( \omega \) and the hypothesis \( \theta \in \omega_i \) restricting the parameters to it.) The hypothesis \( \omega_0 \) itself is assumed to belong to the family and to form the intersection of its members, that is, \( \omega_0 \in \Omega \) and \( \omega_0 = \bigcap_{i \in I} \omega_i \).

The implication relation \( \omega_i \subseteq \omega_j \) for a pair of hypotheses from \( \Omega \) will be referred to by saying \( \omega_j \) is a component of \( \omega_i \). If the relation is strict, i.e., \( \omega_i \subset \omega_j \), \( \omega_j \) will be said to be a proper component of \( \omega_i \). Any such strict implication relation \( \omega_i \subset \omega_j \), where \( i \in I \) and \( j \in I \), may be written \( i < j \) and the set of all such relations existing between pairs of indices from \( I \) will be denoted \( J \). Note that \( i < j \) is a transitive relation, but is neither reflexive nor symmetric and that \( 0 < i \) for all \( i \in I - \{0\} \). Also note that this is not the only type of implication relation that may exist between members of a family of hypotheses. If, for example, \( \mu_1 \), \( \mu_2 \) and \( \mu_3 \) are parameters one might consider the intersection hypothesis \( \omega_0 : \mu_1 = \mu_2 = \mu_3 \) and let the family further contain hypotheses \( \omega_1 : \mu_1 = \mu_2 \), \( \omega_2 : \mu_1 = \mu_3 \) and \( \omega_3 : \mu_2 = \mu_3 \). In this example the relation set \( J \) contains only the relations \( 0 < 1 \) \( 0 < 2 \), \( 0 < 3 \), whereas implications \( (\omega_1 \cap \omega_2 \cap \omega_3) \subset \omega_0 \) or \( (\omega_1 \cap \omega_2) \subset \omega_3 \) are not included in \( J \).

A partial ordering of the hypotheses of \( \Omega \) may be obtained by always assigning a higher rank to a hypothesis \( \omega_i \) than to any of its proper components \( \omega_j \), \( i < j \). This will assign maximum rank to the intersection hypothesis \( \omega_0 \) and minimum rank to hypotheses which have no proper components in \( \Omega \). The latter hypotheses are referred to as minimal and their subfamily denoted \( \Omega_{\text{min}} \) and indexed by \( I_{\text{min}} \) so that \( \Omega_{\text{min}} = \{ \omega_j | j \in I_{\text{min}} \} \). All other hypotheses are referred to as non-minimal, i.e., \( \Omega - \Omega_{\text{min}} = \{ \omega_i | i \in I - I_{\text{min}} \} \).

Corresponding to each hypothesis \( \omega_i \) of \( \Omega \) let there be a real valued statistic \( Z_i = Z_i(Y) \) and write the family of these statistics \( Z = \{ Z_i | i \in I \} \). The collection \( \{ \Omega, Z \} \) of hypotheses and their corresponding statistics will be called a testing family provided the distribution of \( Z_i \), for every \( i \in I \), is completely specified under \( \omega_i \), i.e., the same for all \( \theta \in \omega_i \). If the family \( \Omega \) is closed under intersection the testing family will be called closed. If, for any subfamily \( \tilde{\Omega} = \{ \omega_i | i \in \tilde{I} \} \), where \( \tilde{I} \subseteq I \), the joint distribution of all \( Z_i, i \in \tilde{I} \), is completely specified under
\[ \tilde{\omega}_0 = \bigcap \tilde{\omega}_i, \] the testing family will be called joint. These two categories of testing families are neither exclusive nor exhaustive.

**Example 2.1.** Consider a one-way ANOVA with \( \omega_0 \) as the hypothesis of equality of all \( k (> 3) \) means and \( \omega_i \) the hypothesis of equality of a pair of means indexed by \( I^2 \) \( \equiv \{0\} \) indexing all \( \binom{n}{2} \) such pairs. Note that these \( \binom{n}{2} \) \( \omega_i \)'s are minimal hypotheses, only \( \omega_0 \) being non-minimal. This family is not closed under intersection for it does not contain hypotheses on equality of subsets of \( h \) means if \( 2 < h < k \). However, with \( F \)-ratio statistics (with a common overall estimate of error variance) being used for each hypothesis, a joint testing family \( \{\Omega^F, Z^F\} \) results. For the joint distribution of all the pairwise \( F \)-ratios is completely specified for any subset of \( h(2 \leq h \leq k) \) equal means, irrespective of the values of the other \( k - h \) means. The same obviously also holds for augmented \( F \)-ratios, i.e., \( F \)-ratios multiplied by their numerator df's.

**Example 2.2.** In the same set-up let \( \omega_i \) be a hypothesis of equality for some subset \( S_i \) of \( k \) means, \( I^s \) indexing all \( \binom{k}{2} - k - 2 \) such subsets. The minimal sub-family remains as in Example 2.1. If (augmented) \( F \)-ratio statistics are used for each of these hypotheses, any one of them corresponding to a true hypothesis will have a central (augmented) \( F \) distribution irrespective of any means not involved in that hypothesis. Thus one obtains a closed testing family \( \{\Omega^s, Z^s\} \).

**Example 2.3.** Again, in the same set-up, one may start by ranking the observations of all samples together (overall ranking) and for each hypothesis use the "between" sum of squares of these ranks as a non-parametric statistic. The overall sum of squares is, apart from a constant, the Kruskal-Wallis statistic and its distribution is specified under \( \omega_0 \). But for any subset of \( h(< k) \) samples, the corresponding statistic depends not only on the equality of the \( h \) means involved but also on where the other \( k - h \) means are relative to those \( h \). This is evident when one considers that the ranks available for that subset of \( h \) samples depend on what ranks have been assigned to the other \( k - h \) samples. Thus, such a collection of hypotheses and rank statistics does not form a testing family.

**Example 2.4.** Exactly the same argument holds if instead of "between" sums of squares of ranks one uses any other statistics based on the overall ranking, such as the range of rank means or the maximum rank mean. These types of statistics do not provide testing families. (The use of such statistics and their distributions have been discussed by Nemenyi [17]).

**Example 2.5.** An alternative non-parametric approach to the set-up of Example 2.2 is to use a separate ranking for the statistic of each hypothesis, including only the samples relevant to that hypothesis. One may then use the Kruskal-Wallis statistic for each hypothesis, and as its distribution is specified under the hypothesis a closed testing family is obtained.

Since for each \( \omega_i \) of \( \Omega \) more than one statistic can be chosen to satisfy these conditions, more than one testing family can be defined for any given family of hypotheses. Consideration of the properties of different \( Z_i \)'s as test statistics for their respective hypotheses \( \omega_i \) will play a part in choosing a \( Z \) appropriate to \( \Omega \).
In the present discussion the choice of $Z$ corresponding to any given $\Omega$ is severely restricted by the following requirement of monotonicity.

A testing family $\{\Omega, Z\}$ will be called *monotone* if, whenever $i < j$, the numerical relation

\[(2.1) \quad Z_i(y) \geq Z_j(y)\]

holds a.e. (almost everywhere with respect to measure $\mu$), between the corresponding statistics. If, furthermore, for any non-minimal $\omega_i$, i.e., $i \in I - I_{\min}$, and for any sample point $y$, there exists a proper component $\omega_j$ of $\omega_i$, i.e., $i < j$, for which (2.1) is an equality, that is, if

\[(2.2) \quad Z_i(y) = \max \{Z_j(y) | i < j\} \quad \text{a.e.,}\]

$\{\Omega, Z\}$ will be called *strictly monotone*. Note that the actual $j$ for which $Z_i(y) = Z_i(y)$ need not be the same for each $y$. Clearly, if a testing class is strictly monotone it is also monotone. Also, it is readily seen that if, and only if, (2.2) holds for every $i \in I - I_{\min}$, then, for every $i \in I - I_{\min}$

\[(2.3) \quad Z_i(y) = \max \{Z_j(y) | i < j, j \in I_{\min}\} \quad \text{a.e.}\]

Hence (2.3) is an alternative definition of strict monotonicity. It is also clear that in a monotone testing family the statistic for $\omega_0$, the intersection hypothesis, is

\[(2.4) \quad Z_0(Y) = \max \{Z_i(Y) | i \in I\},\]

since $\Omega$ has been defined to include $\omega_0$ itself.

**Example 2.6.** In the testing family $\{\Omega^\alpha, Z^\alpha\}$ of Example 2.2 implication relations $< \in$ exist between hypotheses on sets of means and hypotheses on proper subsets of these sets. Monotonicity holds since the “between” sum of squares, and hence the augmented $F$-ratio, for any set is no less than that for any of its subsets [6].

**Example 2.7.** If in the subset ANOVA set-up of Examples 2.2 and 2.6, all sample sizes were equal, studentized range statistics $Z^\alpha$ might be used instead of (augmented) $F$-ratios. In that case monotonicity would be strict, for the range of all means always equals the range of some subset of the means.

**Example 2.8.** In the testing family of Example 2.5 it may happen that the Kruskal-Wallis statistic for a subset exceeds that for a set containing that subset. For example, let there be three samples of four observations with the following ranks: sample 1—8, 9, 10, 11; sample 2—1, 2, 6, 7; sample 3—3, 4, 5, 12. The K-W statistic for samples 1 and 2 is 5.33, exceeding that for samples 1, 2 and 3 which is 4.77. Hence that testing family is not monotone.

**Example 2.9.** In a contingency table one may consider the family of independence hypotheses for the entire table as well as for all subtables obtained by omitting or amalgamating rows and/or columns. If Pearson’s chi-square statistic is used to test each hypothesis it may happen that the statistic for a table is less
than that for a sub-table, though independence in the former implies independence in the latter [7]. Hence such a testing family would not be monotone.

Monotonicity holds for all testing families using likelihood ratio or union-intersection statistics, as is shown next.

**Lemma A.** If the statistics of a testing family are all likelihood ratio (LR for short) the family is monotone.\(^2\)

**Proof.** Let the family be \(\{\Omega, Z\}\) and \(\omega_i, \omega_j\) hypotheses of \(\Omega\) such that \(i < j\). The LR statistic for \(\omega_i\) is

\[
Z_i = d(\sup_{\omega_i} f / \sup_{\omega_j} f)
\]

where \(d\) is a monotone decreasing function, and the LR statistic \(Z_j\) for \(\omega_j\) is similarly defined.

Now \(i < j\) is equivalent to \(\omega_i \subset \omega_j\) so that \(\sup_{\omega_i} f \leq \sup_{\omega_j} f\). Thus the ratio argument of \(d\) in \(Z_i\) is no greater than that of \(d\) in \(Z_j\). As \(d\) is decreasing, \(Z_i \geq Z_j\), as was to be proved.

**Example 2.10.** In the ANOVA testing family \(\{\Omega^*, Z\}\) Example 2.2 the F-ratios are LR statistics. Lemma A confirms the monotonicity noted in Example 2.6.

**Example 2.11.** In the contingency table independence testing set-up described in Example 2.9 a monotone testing family may be obtained by using the log likelihood ratio statistics instead of Pearson’s chi-squares [7].

**Lemma B.** A testing family is strictly monotone if and only if its statistics are related by Roy’s Union-Intersection (UI for short) principle [20].

**Proof.** Roy used the UI principle to construct a test for \(\omega_i\) from tests of all its minimal components \(\omega_j, i < j, j \in I_{\min}\). If the latter are tested with critical regions \((y \mid Z_j > \gamma)\), Roy [9], [20] defined the critical region for \(\omega_i, i \in I - I_{\min}\), as \(\bigcup_{j \in I, i < j, j \in I_{\min}} (y \mid Z_j > \gamma)\), or equivalently as

\[
(y \mid \max_j \{Z_j \mid i < j, j \in I_{\min}\} > \gamma).
\]

Extending this definition to all possible values of \(\gamma\), Roy essentially defined the UI statistic for \(\omega_i\) as \(\max_j \{Z_j \mid i < j, j \in I_{\min}\}\). Use of this definition for every \(i \in I\) is equivalent to using the statistics (2.3) which satisfy strict monotonicity, as was to be proved.

**Example 2.12.** For the ANOVA set-up of (2.7) one notes that the studentized range \(Z^*\) for any subset of means is the maximum of the Student \(t\) statistics times \(Z^*, \) i.e., studentized ranges, for all pairwise comparisons of means from that subset. Thus, the studentized range statistics generate UI relations for the testing family \(\{\Omega^*, Z^*\}\) as was expected by Lemma B from the fact that this family is strictly monotone.

Since the UI relation is one of hypotheses and statistics it is preferable not to speak of UI statistics, but of UI related testing families.

3. **Simultaneous test procedures.** For a testing family \(\{\Omega, Z\}\) and a critical
value $\xi$ a Simultaneous Test Procedure (STP) is defined as the family of tests of all $\omega_i \in \Omega$ which reject any $\omega_i$, $i \in I$, if $Z_i > \xi$, using the same constant $\xi$ for all $Z_i$ of $Z$. Such an STP will be denoted $\{\Omega, Z, \xi\}$. The probability
\begin{equation}
\alpha = P_{\omega_0}(Z_0 > \xi)
\end{equation}
of falsely rejecting the intersection hypothesis is referred to as the level of the STP. If $Z_0$ is a continuous variable the same STP may alternatively be defined in terms of its testing family $\{\Omega, Z\}$ and level $\alpha$, the critical value $\xi$ being then determined by (3.1).

The term simultaneous is used to indicate that no order or sequence is imposed on the tests of all the hypotheses of $\Omega$ and that they are made without reference to one another. It is shown below that this cannot lead to incoherent decisions.

Note that $\omega_0$ is rejected by the STP if and only if $Z_0 > \xi$, exactly as by a significance test of significance level $\alpha$. This test of the intersection hypothesis is therefore always part of the STP. If the testing family contains only a single hypothesis $\omega_0$ and its statistic $Z_0$, the STP $\{\Omega, Z, \xi\}$ is equivalent to the test of hypothesis $\omega_0$ by means of statistic $Z_0$ at critical value $\xi$.

An essential requirement that any procedure must satisfy is that no hypothesis should be "accepted" if any hypothesis implied by it is rejected. In other words, if $i < j$ then $\omega_i$ must always be rejected if $\omega_j$ is. This requirement will be called coherence, and an STP will be called coherent if its decisions always satisfy this.

**Theorem 1.** All STP's based on testing family $\{\Omega, Z\}$ are coherent if and only if $\{\Omega, Z\}$ is monotone.

**Proof.** Consider any hypotheses $\omega_i$ and $\omega_j$ of $\Omega$ for which $i < j$. Monotonicity of $\{\Omega, Z\}$ postulates that $Z_i \geq Z_j$ a.e., whereas coherence of $\{\Omega, Z, \xi\}$ for any given $\xi$ requires that $(y | Z_i > \xi) \supseteq (y | Z_j > \xi)$ a.e.$^4$ since these events determine the rejection of $\omega_i$ and $\omega_j$, respectively

Now, the statements "$Z_i \geq Z_j$ a.e." and "$(y | Z_i > \xi) \supseteq (y | Z_j > \xi)$ a.e. for all $\xi$" are clearly equivalent. Hence monotonicity of testing family is equivalent to coherence of all STP's based on that family, as was to be proved.

**Corollary 1.** If the statistics of a testing family are all LR, all STP's based on that family are coherent.

**Proof.** This follows from Lemma A and Theorem 1.

**Example 3.1.** In the subset ANOVA set-up $\{\Omega^8, Z^p\}$ of Examples 2.2, 2.6, 2.10 one may test equality of the $h$ means of each subset by means of the (augmented) $F$-ratio against $\xi^*$, the upper $\alpha$ point of the (augmented) $F$ distribution with $k - 1$ and $n_*$ degrees of freedom, where $n_*$ is the d.f. for the variance estimate. This will give an STP with coherent decisions; set $S_i$ of $k_i$ means will be rejected if any subset of its means is rejected $[6]$.

---

$^4$ In an earlier paper $[6]$ this property was called transitivity. The terminology has been changed because of the different connotations of the earlier term.

$^4$ For set containment relations $A \subset B$ and $A \subseteq B$ the a.e. qualifications mean that $A \cap B$ may be non-empty but of measure zero. Similarly, $A = B$ a.e. mean that both $A \cap B$ and $A \cap B$ may be non-empty but are of measure zero.
Example 3.2. The non-parametric analogue of the above, using the Kruskal-Wallis statistics of Example 2.5, is not an STP since its testing family is not monotone. As was pointed out in Example 2.8, the Kruskal-Wallis statistic for a set may be less than the same statistic for a subset contained in the set. If the common critical value happens to be between these two statistics' values the subset would be rejected whilst the set was accepted. This would lead to the incoherence of asserting that all means of a set are equal whilst some of them differ from each other. Hence, such a procedure would be incoherent. (See further remarks in Section 6, below.)

Example 3.3. In the one-way ANOVA set-up denote the expectations \( \mathbf{y'} = (\mu_1, \mu_2, \ldots, \mu_k) \) and write \( c_i'y_i \in V(C') \) for a contrast in the expectations where \( V(C') \) is the \( k \)-vector space orthogonal to the vector \( \mathbf{1'} = (1, 1, \ldots, 1) \). Now consider \( \omega_j : \mu_1 = \mu_2 = \cdots = \mu_k \) and \( \omega_j : c_i'y_i = \sum_{h=1}^{k} c_{jh}h = 0, j \in I^c - \{0\} \) ranging over all contrasts such that \( c_i'y_i \in V(C') \). Clearly, \( \omega_0 = \bigcap_I \omega_j \) as nullity of all contrasts is equivalent to equality of all expectations.

Let the sample means be written \( \bar{x'} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k) \), the variance estimate \( s^2 \), and the sample sizes \( n_1, n_2, \ldots, n_k \). Then LR statistics are, for \( \omega_0 \), the augmented F-ratio

\[
Z_0 = \left( \frac{\sum_{h=1}^{k} n_h\bar{x}_h^2 - (\sum_{h=1}^{k} n_h\bar{x}_h)^2}{\sum_{h=1}^{k} n_h} \right)/s^2,
\]

and for \( \omega_j, j \in I^c - \{0\} \), the squared Student \( t \)

\[
Z_j = \left( \frac{\sum_{h=1}^{k} c_{jh}\bar{x}_h^2}{s^2} \right)/\sum_{h=1}^{k} n_h (c_{jh}/n_h).
\]

It is readily seen that this provides a joint testing family \( \{\Omega^c, Z^p\} \).

An \( \alpha \) level STP is obtained by choosing \( \tau^p \) as in Example 3.1. This STP consists of testing the overall \( \omega_0 \) and each contrast \( \omega_j \) by checking if \( Z_0 > \tau^p \) and \( Z_j > \tau^p \), respectively. By Corollary 1 this STP \( \{\Omega^c, Z^p, \tau^p\} \) is known to be coherent in the sense that rejection of any contrast hypothesis always entails rejection of the overall equality hypothesis.

It will be noted that this STP consists of the significance tests of Scheffé's method (Example 5.1 [25, 6]).

In a coherent STP the level \( \alpha \) may be regarded as an experimentwise level, as follows from the next theorem.

Theorem 2. The probability that a coherent STP \( \{\Omega, Z, \tau\} \) of level \( \alpha \) rejects at least one true \( \omega_j \) of \( \Omega \) is \( \alpha \) if \( \omega_0 \) is true; it is at most \( \alpha \) irrespective of the truth of \( \omega_0 \) provided \( \{\Omega, Z\} \) is either closed or joint.

The probability of rejecting any particular true \( \omega_i \) of \( \Omega \) is no more than the above probability.

Proof. By Theorem 1, \( \{\Omega, Z\} \) must be monotone, so that by (2.4) \( Z_0 = \max \{Z_i : i \in I\} \). Hence rejection of \( \omega_0 \), event \( (Z_0 > \tau) \), is equivalent to rejection of any \( \omega_j \), event \( U_1(Z_i > \tau) \). Under \( \omega_0 \) the probability of the former event is \( \alpha \), (3.1), hence so is the probability of the latter, proving the first statement.

Next, let \( \bar{\Omega} = \{\omega_i : i \in \bar{I}\}, \bar{I} \subseteq I \), be the subfamily of true hypotheses whose intersection \( \bar{\omega}_0 = \bigcap_{\bar{I}} \omega_i \) is necessarily also true. Then rejection of a true \( \omega_i \) occurs
whenever $\max \{ Z_i | i \in I \} > \xi$ and the required probability is

$$P_{\omega_0}(\max \{ Z_i | i \in I \} > \xi).$$

If it can be shown that

$$(3.2) \quad P_{\omega_0}(\max \{ Z_i | i \in I \} > \xi) = P_{\omega_0}(\max \{ Z_i | i \in \bar{I} \} > \xi)$$

then the second statement of the theorem follows, for the latter probability is clearly no greater than $P_{\omega_0}(\max \{ Z_i | i \in I \} > \xi) = P_{\omega_0}(Z_0 > \xi) = \alpha$.

If $\{ \Omega, Z \}$ is closed $\omega_0 \in \Omega$, as $\Omega$ is closed under intersection. Hence max $\{ Z_i | i \in I \}$, the statistic for $\omega_0$, has its distribution completely specified under $\omega_0$, irrespective of whether $\omega_0$ holds or not, and so $(3.2)$ holds.$^6$

If $\{ \Omega, Z \}$ is joint, the joint distribution of all $Z_i$, $i \in I$, is specified under $\omega_0$, irrespective of the truth of $\omega_0$. Hence the same holds also for $\max \{ Z_i | i \in I \}$ and $(3.2)$ follows as before.

The last statement in the theorem is obvious.

Coherence prevents the contradiction of rejecting a hypothesis without also rejecting all other hypotheses implying it. It does not, however, preclude the dissonance of rejecting a hypothesis whilst not rejecting any other hypotheses implied by it. Such dissonances, though not desirable, are sometimes allowed when the alternative to rejection of a hypothesis is taken to be non-rejection rather than acceptance. They are familiar to users of tests of significance, as these merely reject a hypothesis without indicating where it fails. They occur less often when STP's are used, as they may indicate some components to be also rejected. However, the use of STP's does not eliminate all dissonances: some STP's may reject hypotheses without rejecting any of their proper components—such STP's will be called non-consonant. One would generally prefer procedures which are coherent as well as consonant. In a later section non-consonant procedures will be compared in terms of the likelihood of avoiding dissonances.

Note that consonance of an STP can equivalently be defined as the requirement that any $\omega_i$ of $\Omega$ be rejected only if some minimal component of $\omega_i$ is rejected.

**Example 3.4.** In the subset ANOVA F-ratio STP $\{ \Omega^g, Z^F, \xi^F \}$ of Example 3.1 the minimal hypotheses $\omega_j : \mu_{ja} = \mu_{jb}, j \in I_{\min}^g$, are those of pairwise equality of means. Since

$$Z_0^F = (\sum_{h=1}^k n_h \bar{x}_h^2 - (\sum_{h=1}^k n_h \bar{x}_h)^2)/\sum_{h=1}^k n_h / s^2$$

and

$$Z_j^F = (\bar{x}_{ja} - \bar{x}_{jb})^2 / s^2(1/n_{ja} + 1/n_{jb}), \quad j \in I_{\min}^g,$$

and

$$Z_0^F > \max \{ Z_j^F | j \in I_{\min}^g \} \quad \text{a.e.} \quad \text{\cite{Note}}$$

$^6$ A similar proof for one-way ANOVA was given in [6] and a general proof for closed testing families in [11] by Knight. As a result of the requirement of closure under intersection Knight's form of the theorem cannot be applied directly to some important classes of procedures, as, for instance, Scheffé's (Example 3.3, above). Knight surmounts this difficulty by applying his lemma to an extended family which includes all intersections of sets from $\Omega$. 


(see [9]), \( (Z_0 \mathbf{Z}^x > \xi) \) does not necessarily imply \( (Z_j \mathbf{Z}^x > \xi) \) for any \( j \in I_{\text{min}}^s \). In other words, dissonances may occur in rejecting over-all equality of means without rejecting the equality of any pair of means [6].

**Example 3.5.** The contrast ANOVA \( F \)-ratio STP \( \{ \Omega^c, \mathbf{Z}^x, \xi^r \} \) of Example 3.3 is consonant as well as coherent. To see this, note that the maximal squared \( t \) statistic for a contrast is equal to the overall augmented \( F \)-ratio, i.e.,

\[
Z_0^r = \max \{ Z_j^r \mid j \in I_{\text{min}}^c \}
\]

where \( I_{\text{min}}^c = I^c - \{0\} \). (This is established by differentiating the \( Z_j^r \) expression in Example 3.3 over all \( c_j \) and setting the derivative equal to zero [6]. See also [26]). Hence, \( (Z_0^r > \xi^r) = U_{I_{\text{min}}^c} (Z_j^r > \xi^r) \), so that whenever overall \( \omega_0 \) is rejected so is \( \omega_j \) for some contrast.

**Example 3.6.** In the pairwise ANOVA set-up the \( F \)-ratio STP \( \{ \Omega^p, \mathbf{Z}^p, \xi^p \} \) would consist of tests belonging to both \( \{ \Omega^s, \mathbf{Z}^s, \xi^p \} \) of Example 3.1 and \( \{ \Omega^c, \mathbf{Z}^c, \xi^p \} \) of Example 3.3, since the overall \( \omega_0 \) is the same in \( \Omega^p \), \( \Omega^s \) and \( \Omega^c \) and the other hypotheses of \( \mathbf{Z}^p \), that is, those on pairwise comparisons, are included in both \( \Omega^p \) and in \( \Omega^c \).

**Theorem 3.** All STP’s based on testing family \( \{ \Omega, \mathbf{Z} \} \) are coherent and consonant if and only if \( \{ \Omega, \mathbf{Z} \} \) is strictly monotone.

**Proof.** STP \( \{ \Omega, \mathbf{Z}, \xi \} \) is coherent and consonant, by definition, if, for every \( i \in I \),

\[
(y \mid Z_i > \xi) = U_{i < i, i \in I_{\text{min}}} (y \mid Z_j > \xi) \quad \text{a.e.}
\]

For coherence requires the right hand side event to imply the left hand side, whereas consonance requires that the left hand side imply the right hand side.

All STP’s based on \( \{ \Omega, \mathbf{Z} \} \) are coherent and consonant if and only if, (3.3) hold for all \( \xi \), but this is equivalent to (2.3), the definition of strict monotonicity of that testing family.

**Corollary 2.** If, and only if, a testing family is UI related, all STP’s based on that family are coherent and consonant.

**Proof.** This follows from Lemma B and Theorem 3.

Corollary 2 points out the correspondence between UI statistics and consonant STP’s. Roy has used the UI principle to induce a test of the intersection hypothesis from the tests of its components, whereas its equivalent is used here to resolve the test decision on the intersection hypothesis into decisions on the components. The UI test of \( \omega_0 \) can be used as part of an STP. Rejection by this test is followed in the STP by further tests to resolve which components are to be rejected, whereas non-rejection obviates the need for further testing.

**Example 3.7.** In the subset ANOVA set-up for equal sample sizes \( n_1 = n_2 = \cdots = n_k = n \), say, one might use studentized range statistics as in Example 2.7. Thus, hypothesis \( \omega_i : \mu_{i_1} = \mu_{i_2} = \cdots = \mu_{i_k} \) of equality of all \( k \) means of subset \( S_i \), say, would be tested with statistic

\[
Z_i^r = \max \{|\bar{x}_{i_1} - \bar{x}_{i_2}| \mid i_a \neq i_b \in S_i \}.
\]

Taking \( \xi^r \) as the upper \( \alpha \) point of the studentized range distribution for \( k \) variables with \( n \) d.f., and rejecting any \( \omega_i \) if \( (Z_i^r > \xi^r) \), one obtains an STP
\{\Omega^k, Z^k, \xi^k\}. As pointed out in Example 2.12 this testing family \{\Omega^k, Z^k\} is UI related and hence is strictly monotone. By Corollary 2 \{\Omega^k, Z^k, \xi^k\} is coherent and consonant.

**Example 3.8.** If the decisions of the STP of Example 3.7 were restricted to the overall hypothesis and all pairwise comparisons (as in Example 3.6) this would not affect the coherence and consonance properties and one would obtain an STP \{\Omega^k, Z^k, \xi^k\} consisting of the significance tests of Tukey’s method of allowances [27].

Coherence and consonance are relations between test decisions on hypotheses which imply one another, i.e., between some \(\omega_i\) and \(\omega_j\) of \(\Omega\) for which \(i < j\). As was noticed in Section 2, above, other implication relations may also obtain within the family \(\omega\). The STP properties of coherence and consonance do not relate to such other relations and indeed STP decisions may violate them. Thus, in a family of hypotheses on equality of sets of parameters from among \(\mu_1, \mu_2, \ldots, \mu_k\), an STP may accept \(\mu_1 = \mu_2\) as well as \(\mu_1 = \mu_3\) but reject \(\mu_2 = \mu_3\) even though it is implied jointly by the two accepted hypotheses. Lehmann [14], [15] has considered procedures which preserve a wider class of implication relations and has termed them compatible. STP’s, even consonant ones, are not generally compatible in that sense.

**4. Resolution of STP’s.** The purpose of multiple comparisons in general, and STP’s in particular, is to provide resolution of significant test decisions on overall hypotheses into significant decisions on components, as far down as minimal components. Thus, for family \(\Omega = \{\omega_i | i \in I\}\) rejection of intersection \(\omega_i\) is to be resolved into rejection of proper components \(\omega_i, i \in I,\) down to minimal hypotheses \(\omega_j, j \in I_{\text{min}}\).

The extent of resolution of a procedure must be defined in terms of the likelihood of rejecting minimal proper components provided the overall hypothesis is rejected. If two procedures each test overall hypothesis \(\omega_0\) at level \(\alpha\), the one will be said to be no less resolvent than the other if it always rejects any minimal hypothesis rejected by the other, and will be said to be strictly more resolvent if it sometimes rejects when the other does not. In other words, consider two STP’s \{\Omega, Z, \xi\} and \{\Omega^*, Z^*, \xi^*\}, such that \(\omega_0 = \bigcap_I \omega_i = \bigcap_I^r \omega_i\), where \(\Omega = \{\omega_i | i \in I\}\) and \(\Omega^* = \{\omega_i | i \in I^*\}\) and \(I_{\text{min}} \subset I^*\). Then the former STP is no less resolvent than the latter if

\[
(4.1a) \quad P_{\omega_0}(Z_0 > \xi) = P_{\omega_0}(Z_0^* > \xi^*)
\]

and for all \(j \in I_{\text{min}}\)

\[
(4.1b) \quad (Z_j > \xi) \supseteq (Z_j^* > \xi^*) \quad \text{a.e.}
\]

It is strictly more resolvent if the containment in (4.1b) is proper. It is clear that no meaningful comparisons can be made between STP’s unless they test the same hypotheses, at least to the extent that their intersections are the same and the minimal hypotheses of one of the families are all included in the other family.
**Example 4.1.** In the pairwise ANOVA set-up with equal sample sizes the studentized range STP $\{\Omega^*, Z^*, \zeta^*\}$ (Tukey's method—Example 3.8) will reject any pairwise comparison of means which is rejected by the $F$-ratio STP $\{\Omega^o, Z^o, \zeta^o\}$ (Scheffé's method—Example 3.5) provided both STP's are of the same level [9]. Note that the family of hypotheses of the latter STP is wider than that of the former as it also includes all contrasts, but that all hypotheses of the former are included in the latter family.

Resolution comparisons are made between STP's of the same level. An alternative kind of comparison can be made between STP's which give identical decisions on all minimal components. The procedure will be said to be no less (more) parsimonious if it gives these decisions at no more (less) cost in rejection of other, non-minimal, hypotheses, and specifically of the overall hypothesis. Thus $\{\Omega, Z, \zeta\}$ is no less parsimonious than $\{\Omega^*, Z^*, \zeta^*\}$, where $\Omega \subseteq \Omega^*$ and $\omega_0 = \omega_0^*$, if

\begin{align}
(4.2a) & \quad \text{for all } j \in I_{\text{min}} \quad (Z_j > \zeta) = (Z_j^* > \zeta^*) \quad \text{a.e.}, \\
(4.2b) & \quad \text{for all } i \in I \quad (Z_i > \zeta) \subseteq (Z_i^* > \zeta^*) \quad \text{a.e.}, \\
(4.2c) & \quad \text{and for } \omega_0 \quad P_{\omega_0}(Z_0 > \zeta) \leq P_{\omega_0}(Z_0^* > \zeta^*).
\end{align}

If the inequality in (4.2c) is strict, $\{\Omega, Z, \zeta\}$ is strictly more parsimonious than $\{\Omega^*, Z^*, \zeta^*\}$.

A possible measure of resolution is the ratio of the rejection probabilities of minimal hypotheses to that of the overall hypothesis. Under the overall hypothesis $\omega_0$, if all minimal hypotheses have the same probability of being rejected, i.e., if

$$\alpha_1 = P_{\omega_0^*}(Z_j > \zeta)$$

for all $j \in I_{\text{min}}$, resolution may be gauged by the ratio $\alpha_1/\alpha$. Clearly, if one STP is more resolvent than another, the former will have a higher $\alpha_1/\alpha$ ratio than the latter. However, a higher $\alpha_1/\alpha$ ratio merely means more probable resolution, not necessarily strictly more resolution in the sense of (4.1b).

**Example 4.2.** Consider an eight mean ANOVA with 40 d.f. for error [6]. The 10% level studentized range STP $\{\Omega^s, Z^s, 4.10\}$ has a resolution ratio of $\alpha_1/\alpha = 0.006/0.100 = 0.06$ whereas the 10% $F$-ratio STP $\{\Omega^o, Z^o, 13.09\}$, noted in Example 4.1 to be less resolvent, has a ratio of only $\alpha_1/\alpha = 0.001/0.100 = 0.01$.

**Example 4.3.** In the subset ANOVA set-up any hypothesis $\omega_i$ on a subset $S_i$ of $k_i(>2)$ expectations may be tested by means of the maximal pairwise $F$ statistic

$$Z_i^M = \max_{s_j \subseteq S_i} Z_j^s,$$

where the index $j$ indicates $S_j$ is a pair. For equal sample sizes $Z_j^s = (Z_j^s)^2/2$ (Example 2.12) so that the use of $Z_i^M$ is equivalent to the use of the studentized range (Example 3.7). For unequal samples percentage points of $Z_0^M$ are not available. For any critical value $\zeta^M$, the STP $\{\Omega^s, Z^M, \zeta^M\}$ is clearly coherent and consonant but its level $\alpha$ cannot be determined exactly for unequal $n_j$s. However, by virtue of the first Bonferroni inequality $\alpha \leq (\frac{k}{2})\alpha_1$, where $\alpha_1$ is the pair-
wise probability of a type I error which is readily determined by noting that $\xi^*$ is the upper $\alpha_1$ point of the $F$ distribution with 1 and $n_e$ d.f. The resolution ratio is thus $\alpha_1/\alpha \geq 1/(1/2)$ for the maximal pairwise $F$ STP. For the particular case of $k = 8$, $\alpha_1/\alpha \geq 0.036$, whereas, with $n_e = 40$, for the augmented $F$-ratio STP this ratio is 0.01 (as in Example 4.2). Thus, if tests are desired only for pairs and other sets, and not for all contrasts, one would have more probable resolution with this approximate STP using the maximal pairwise $F$-ratio than with the one using the ordinary $F$-ratio, and one may well be led to prefer to use the former.

A more general measure of resolution should also take into account the chances of rejecting non-minimal component hypotheses. If ranks can be assigned meaningfully to hypotheses, resolution should depend on the entire sequence

$$\alpha_1 : \alpha_2 : \cdots : \alpha_{k-1} = \alpha$$

where $\alpha_q$ is the probability of falsely rejecting a hypothesis of rank $q$ and $k - 1$ is the maximum rank (as in a $k$ mean ANOVA), i.e., the rank of the intersection hypothesis. Examples of such sequences of probabilities are given in [6] for ANOVA STP's and in [7] for contingency table STP's.

To determine which STP's are more resolvent or more parsimonious than which others, one is led to consider the following relation between monotone testing families. $\{\Omega, Z\}$ will be said to be narrower than $\{\Omega^*, Z^*\}$ if $\Omega \subseteq \Omega^*$ and $\omega_0 = \omega_0^*$ and if there exists a function $g$ such that

\begin{align*}
(4.3a) & \quad g \text{ is continuous monotone increasing,} \\
(4.3b) & \quad \text{for every } j \in I_{\min} \quad Z_j^* = g(Z_j) \quad \text{a.e.,} \\
(4.3c) & \quad \text{for every } i \in I - I_{\min} \quad Z_i^* \geq g(Z_i) \quad \text{a.e.,}
\end{align*}

and

\begin{align*}
(4.3d) & \quad P(Z_0^* > g(Z_0)) > 0.
\end{align*}

A stronger condition which implies (4.3d) is

\begin{align*}
(4.3d') & \quad Z_0^* > g(Z_0) \quad \text{a.e.}
\end{align*}

If (4.3d') holds $\Omega, Z$ will be said to be strictly narrower than $\Omega^*, Z^*$.

**Example 4.4.** Again, in the subset ANOVA set-up for equally sized samples, compare the studentized range testing family $\{\Omega^s, Z^s\}$ with the augmented $F$-ratio testing family $\{\Omega^s, Z^s\}$. The hypothesis family $\Omega^s$ and intersection $\omega_0 = \mu_1 = \mu_2 = \cdots = \mu_k$ are the same for both. Any minimal hypothesis $j \in I_{\min}^s$ is a pairwise comparison of expectations $\omega_j : \mu_{j_a} - \mu_{j_b} = 0$, for which $Z_j^* = (Z_j^s)^2 / 2$ (Example 2.12). Since $Z^s$ can assume only positive values the function $Z^s = (Z^s)^2 / 2$ satisfies (4.3a) and (4.3b). Also, it has been shown ([9], see also Example 4.10, below) that for all non-minimal hypotheses $\omega_i$ on $k_i (> 2)$ expectations $Z_i^s > (Z_i^s)^2 / 2$ a.e., so that (4.3c) and (4.3d') also hold. This shows $\Omega^s, Z^s$ to be narrower than $\Omega^s, Z^s$.
EXAMPLE 4.5. For contrast hypotheses \( \omega_j : c_j' y = \sum_{h=1}^{k} c_{jh} \mu_h = 0 \) Tukey has generalized the studentized range statistic as
\[
Z_j^R = 2n^{1/2} \left| \sum_{h=1}^{k} c_{jh} \bar{\xi}_h \right| / \sqrt{\sum_{h=1}^{k} |c_{jh}|}
\]
and shown that \( Z_0^R = \max \{ Z_j^R | c_j \in V(C') \} \) (26), Section 3.6, [28], [16] pp. 44–46). Hence \( \{ \Omega^C, Z^R \} \) is UI related, just as \( \{ \Omega^C, Z^R \} \) is (Example 3.5 and Corollary 2). Thus, in the contrast ANOVA set-up for equally sized samples a generalized studentized range STP \( \{ \Omega^C, Z^R, \xi^R \} \) is an alternative to the (augmented) F-ratio STP \( \{ \Omega^C, Z^R, \xi^R \} \) of Example 3.5.

Let \( g(x) = x^2 / 2 \) for \( x > 0 \). This is a function which satisfies (4.3a), (4.3b) for all pairwise contrasts and (4.3c) and (4.3d) for the non-minimal hypothesis \( \omega_0 : \mu_1 = \mu_2 = \cdots = \mu_k \) (Example 4.3). However, for minimal hypotheses on contrasts involving more than two expectations (4.3b) does not hold. Hence \( \{ \Omega^C, Z^R \} \) is not narrower than \( \{ \Omega^C, Z^R \} \), nor vice versa.

EXAMPLE 4.6. Also compare testing families \( \{ \Omega^*, Z^R \} \) of Example 2.1 and \( \{ \Omega^C, Z^R \} \) of Example 3.3. Clearly, \( \Omega^* \subset \Omega^C \) and \( \omega_0 \) is the same for both. Furthermore, as in Example 4.3, for all \( j \in I_{\min}^P \)
\[
Z_j^P = (Z_j^R)^2 / 2 \quad \text{and} \quad Z_0^P > (Z_0^R)^2 / 2 \quad \text{a.e.,}
\]
\( \omega_0 \) being the only non-minimal hypothesis of \( \Omega^* \) (and \( \Omega^C \)). Hence (4.3a, b, c, d) hold so that \( \{ \Omega^*, Z^R \} \) is found to be narrower than \( \{ \Omega^C, Z^R \} \).

Given the narrowness relation between testing families the following two theorems show the corresponding relations between STP’s based on these families.

THEOREM 4. If \( \{ \Omega, Z \} \) is narrower than \( \{ \Omega^*, Z^* \} \), then for every \( \xi^* \) there exists \( \xi \) such that \( \{ \Omega, Z, \xi \} \) is no less parsimonious than \( \{ \Omega^*, Z^*, \xi^* \} \). Moreover, the former STP is strictly more parsimonious than the latter for some values of \( \xi^* \), i.e., those for which
\[
P_{\omega_0} (Z_0^* > \xi^* \geq g(Z_0)) > 0,
\]
where \( g \) satisfies (4.3a, b, c, d).

Proof. In view of narrowness conditions, (4.3a, b, c, d) hold for some function \( g \). For any given \( \xi^* \) choose \( \xi \) so that \( \xi^* = g(\xi) \). Then it follows from (4.3b) that
for all \( j \in I_{\min} \) \( (y | Z_j^*(y) > \xi^*) = (y | Z_j(y) > \xi) \) a.e.,
and from (4.3c) that
for all \( i \in I \) \( (y | Z_i^*(y) > \xi^*) \supseteq (y | Z_i(y) > \xi) \) a.e.,
establishing (4.2a) and (4.2b). From (4.3c) one further obtains
\[
P_{\omega_0} (Z_0^* > \xi^*) = P_{\omega_0} (Z_0 > \xi) + P_{\omega_0} (Z_0 > \xi^* \geq g(Z_0))
\]
from which (4.2c) follows—and no lesser parsimony is proved. It also follows that (4.2c) is a strict inequality—more parsimony—if (4.4) holds. That this
holds at least for some values of $\xi^*$ is ensured by (4.3d), and the theorem is proved.

**Example 4.7.** In comparing the subset ANOVA studentized range STP \[\{\Omega^l, Z^l, \xi^l\}\] and augmented F-ratio STP \[\{\Omega^a, Z^a, \xi^a\}\], choose $\xi^a = (\xi^a)^2/2$. In view of the narrowness established in Example 4.4 it follows from Theorem 4 that the pairwise decisions of both STP's are identical but for all hypotheses on three or more means the F-ratio STP will reject whenever the studentized ratio STP does and more often. In particular, this will result in a higher level for the F-ratio STP—see Table 1 in [6].

**Theorem 5.** If \[\{\Omega, Z\}\] is narrower than \[\{\Omega^a, Z^a\}\] and if \[\{\Omega, Z, \xi\}\] and \[\{\Omega^a, Z^a, \xi^a\}\] are of the same level, then the former STP is no less resolvent than the latter. Moreover, it is strictly more resolvent for some values of \(\xi\), i.e., those for which

\[
P_{w_0}(Z_0^a > g(\xi) \geq g(Z_0)) > 0,
\]

where \(g\) satisfies (4.3a, b, c, d).

**Proof.** According to the proof of Theorem 4, \(P_{w_0}(Z_0^a > g(\xi)) \geq P_{w_0}(Z_0 > \xi)\) for all \(\xi\) with strict inequality if (4.5) holds. Hence, if the levels of the two STP's are to be equal, i.e., if \(P_{w_0}(Z_0^a > \xi^a) = P_{w_0}(Z_0 > \xi)\) (condition (4.1a)), one must have \(\xi^a > g(\xi)\), with strict inequality conditional on (4.5). Now, minimal hypotheses \(\omega_j, j \in I_{min}\), will be rejected by \(\{\Omega^a, Z^a, \xi^a\}\) when \((Z_j^a > \xi^a)\). But, in view of (4.3a)

\[
(Z_j^a > \xi^a) = (g(Z_j) > \xi^a) \subseteq (g(Z_j) > g(\xi)) = (Z_j > \xi),
\]

the containment being proper if (4.5) holds. Thus (4.1b) is established generally and the resolvent of \(\{\Omega, Z, \xi\}\) shown to be always no less, and strictly whenever (4.5) holds. The latter occurs for at least some values of \(\xi\), as argued in Theorem 4 for (4.4).

**Example 4.8.** Again, comparing the pairwise ANOVA studentized range STP \[\{\Omega^l, Z^l, \xi^l\}\] with the contrast ANOVA F-ratio STP \[\{\Omega^c, Z^c, \xi^c\}\], the critical values may be chosen to ensure equal levels. In view of the narrowness comparison of the testing families in Example 4.6, it follows from Theorem 5 that \(\{\Omega^c, Z^c, \xi^c\}\) is more resolvent than \(\{\Omega^a, Z^a, \xi^a\}\), as stated in Example 4.1.

**Example 4.9.** Numerical calculations have shown ([25], [26], Chapter 3) that neither Tukey's nor Scheffe's method is more resolvent for all contrasts (i.e., narrower confidence bounds—see Section 5, below). This accords with the finding in Example 4.5 that neither of the corresponding testing families is narrower than the other.

It is evident from these theorems that narrower testing families have the advantages of providing more⁶ resolvent STP's of the same level (Theorem 5) and more⁶ parsimonious STP's for the same decisions on minimal hypotheses

⁶ "Providing more resolvent and more parsimonious STP's" must be understood as in Theorems 4 and 5 as strictly more for some critical values and no less for all other critical values.
(Theorem 4). In practical applications many multiple comparisons techniques have been criticised precisely for insufficient resolution or excessively frequent overall rejections for given minimal decisions. Hence, the use of narrower testing families may help to provide more practically useful techniques.

The next lemma gives a sufficient condition for one testing family to be narrower than another. In the following corollary this condition is applied to show when some STPs will be more resolvent than others.

**Lemma C.** Let \( \{\Omega, \mathbf{Z}\} \) be a strictly monotone testing family and \( \{\Omega^*, \mathbf{Z}^*\} \) be a monotone testing family such that \( \Omega \subseteq \Omega^* \), \( \Omega_{\text{min}} \subseteq \Omega^*_{\text{min}} \) and \( \omega_0 = \omega_0^* \). A sufficient condition for \( \{\Omega, \mathbf{Z}\} \) to be narrower than \( \{\Omega^*, \mathbf{Z}^*\} \) is the existence of a function \( g \) satisfying (4.3a, b) and also

\[
(4.3e) \quad P_{\omega_0}(Z_0^* \neq g(Z_0)) > 0.
\]

If (4.3e) is replaced by

\[
(4.3e') \quad Z_0^* \neq g(Z_0) \quad \text{a.e.},
\]

one obtains a sufficient condition for strict narrowness.

**Proof.** If (4.3e) holds, (4.3e) implies (4.3d) and thus narrowness, whereas (4.3e') implies (4.3d') and thus strict narrowness. It therefore suffices to establish (4.3e).

Now, for any \( i \in I - I_{\text{min}} \), strict monotonicity of \( \{\Omega, \mathbf{Z}\} \) gives, by (2.3), that

\[
Z_i = \max \{Z_j | i < j, j \in I_{\text{min}}\} \quad \text{a.e.},
\]

whereas monotonicity of \( \{\Omega^*, \mathbf{Z}^*\} \) requires, by (2.1) that

\[
Z_i^* \geq \max \{Z_j^* | i < j, j \in I_{\text{min}}^*\} \quad \text{a.e.}
\]

Under the conditions of this lemma

\[
\max \{Z_j^* | i < j, j \in I_{\text{min}}^*\} \geq \max \{Z_j^* | i < j, j \in I_{\text{min}}\}
\]

\[
= \max \{g(Z_j) | i < j, j \in I_{\text{min}}\}
\]

\[
= g(\max \{Z_j | i < j, j \in I_{\text{min}}\}),
\]

so that, for all \( i \in I - I_{\text{min}} \), \( Z_i^* \geq g(Z_i) \) a.e., which is condition (4.3c).

**Example 4.10.** In the subset ANOVA set-up for equal sample sizes the studentized range testing family \( \{\Omega^s, \mathbf{Z}^s\} \) was noted (Example 2.7) to be strictly monotone and the augmented \( F \)-ratio testing family \( \{\Omega^s, \mathbf{Z}^r\} \) to be monotone (Example 2.6). The family of hypotheses is the same in both cases and the monotonic increasing relation \( Z_i^r = (Z_i^s)^2/2 \) holds for all minimal (pairwise) hypotheses (Example 4.4). For any other hypothesis \( \omega_i \) on \( k_i (>2) \) means, the statistic

\[
Z_i^r = \sum_{h=1}^{k_i} n_{ih} (\bar{x}_{ih} - \bar{x}_i)^2/s^2
\]

where

\[
\bar{x}_i \sum_{h=1}^{k_i} n_{ih} = \sum_{h=1}^{k_i} n_{ih} \bar{x}_{ih},
\]
and the statistic

\[(Z_i^R)^2/2 = \max |\bar{x}_{i_1} - \bar{x}_{i_2}|n^4/s_i^2 | i_1, i_2 \in S_i|\]

are equal only if \(k_i - 2\) of the means \(\bar{x}_{i_1}\) are exactly equal to the average of the other 2 means, clearly an event of probability zero. Hence \(Z_i^* = (Z_i^R)^2/2\) a.e., and all the conditions of Lemma C are seen to obtain. Thus the unproven statement in Example 4.3 is established and the proof that \( \{\Omega^*, Z^R\} \) is narrower than \( \{\Omega^*, Z^R\} \) is complete.

**Corollary 3.** Let \( \{\Omega, Z\} \) and \( \{\Omega^*, Z^*\} \) be monotone testing families such that \( \Omega \subseteq \Omega^*, \Omega_{\min} \subseteq \Omega_{\min}^* \) and \( \omega_0 = \omega_0^* \) and let there exist a function \( g \) satisfying (4.3a, b). Then each of the following are sufficient conditions that \( \{\Omega, Z\} \) provide more parsimonious and more resolvent STP's than \( \{\Omega^*, Z^*\} \),

1. If \( \{\Omega, Z\} \) is UI related and \( g \) also satisfies (4.3e) or (4.3e');
2. If both \( \{\Omega, Z\} \) and \( \{\Omega^*, Z^*\} \) are UI related and

\[
P_{\omega_0}(\max \{Z_j^* | j \in I_{\min}^*\} > \max \{Z_j^* | j \in I_{\min}\} > 0.\]

**Proof.** By Lemma B, both (I) and (II) require \( \{\Omega, Z\} \) to be strictly monotone. Hence, by Lemma C, condition (I) is sufficient for \( \{\Omega, Z\} \) to be narrower than \( \{\Omega^*, Z^*\} \). The rest follows from Theorems 4 and 5.

Under (II), both testing families are strictly monotone so that, a.e.,

\[Z_0^* = \max \{Z_j^* | j \in I_{\min}^*\} \quad \text{and} \quad Z_0 = \max \{Z_j | j \in I_{\min}\}.\]

Now by (4.3a, b)

\[g(Z_0) = \max \{g(Z_j) | j \in I_{\min}\} = \max \{Z_j^* | j \in I_{\min}\},\]

so that (4.6) of (II) is equivalent to (4.3d) and this implies (4.3e). The rest of the argument is as for (I), above.

**Example 4.11.** The conclusions of Example 4.7 regarding equal sample size ANOVA STP's for subsets would also follow from Corollary 3(I) upon noting the existence of a \( g \) function satisfying (4.3a, b, c) as in Example 4.10, and the fact that the studentized range testing family is UI related (Example 2.12).

**Example 4.12.** Similarly, the conclusions of Example 4.8 regarding pairwise and contrastwise STP's follow from Corollary 3 (II). Note first, by Example 4.6, that a function \( x^2/2 \) exists which satisfies (4.3a, b). Next, note that both \( \{\Omega^*, Z^R\} \) and \( \{\Omega^C, Z^C\} \) are UI related (Examples 3.8 and 3.5, respectively).

Finally, condition (II) is established since (4.6) holds, i.e.,

\[
\max \{Z_j^R | c_j \in V(C')\} > \max \{Z_j^R | j \in I_{\min}\} \quad \text{a.e.,}
\]

for otherwise a number of contrasts in means \( \bar{x}_a \) have to be zero, an event of probability zero.

**Example 4.13.** In the equal sample size ANOVA set-up let \( \Omega^D = \{\omega_0, \omega_2, \omega_3, \cdots, \omega_k\} \), where \( \omega_0 : \mu_1 = \mu_2 = \cdots = \mu_k \) as before and \( \omega_j : \mu_1 = \mu_j, j = 2, \cdots, k \) so that

\[\omega_0 = \cap_{j=2}^k \omega_j\]

is the intersection of \( \Omega^D \), and \( \omega_j, j = 2, \cdots, k \), are minimal. Let \( Z_j^D, j = 2, \cdots, k \), be the absolute value of Student's \( t \) for \( \omega_j \), so that \( Z_j^D = (Z_j^D)^2 = (Z_j^R)^2/2 \),
Further, let \( Z_0^D = \max \{ Z_j^D : j = 2, \cdots, k \} \) so that \( \{ \Omega^D, Z^D \} \) is UI related. An \( \alpha \) level STP \( \{ \Omega^D, Z^D, \xi^D \} \) may be obtained by choosing \( \xi^D \) from the tables provided by Dunnett [5].

To compare Dunnett's technique with Tukey's (Example 3.8) note that \( \Omega^D \subset \Omega^R, \Omega^D_{\text{min}} \subset \Omega^R_{\text{min}} \) and \( \omega_0 \) is the intersection of both. The two testing families are UI related, so it remains only to check whether (4.6) holds. This is so since the event \( \max \{ Z_j^R : j \in I_{\text{min}}^R \} > \max \{ Z_j^D : j = 2, \cdots, k \} \) may obviously occur with positive probability (but less than one) under \( \omega_0 \).

It then follows from Corollary 3 (II) that if \( \xi^D = 2^R \xi^R \) the STP \( \{ \Omega^D, Z^D, \xi^D \} \) will have identical decision on \( \omega_j, j = 2, \cdots, k, \) as STP \( \{ \Omega^R, Z^R, \xi^R \} \), but it will be more parsimonious. On the other hand, if \( \xi^D \) and \( \xi^R \) are chosen so that both STP's are of the same level, the former will be more resolvent than the latter.

An analogous result was proved by Krishnaiah for the related confidence statements ([13] and Example 5.9, below).

Corollary 3 allows parsimony and resolution comparisons of different STP's with equivalent statistics for minimal hypotheses. Condition I of the corollary may be used to compare STP's testing the same family of hypotheses (as in Example 4.11) in which case it shows those with UI related testing classes to be preferable to any others. Condition (II) of the corollary may be used to compare STP's all of which have UI related testing classes (Examples 4.12, 4.13). In that case the most preferable STP is shown to be the one whose subfamily of minimal hypotheses is contained in that of the other STP's. This is so unless the containment is trivial in the sense that the maximum of the STP's statistics for the contained subfamily equals that for a containing subfamily with probability one. The importance of using the UI principle with as restricted a family of hypotheses as possible was already stressed by Roy and Srivastava [24]. Krishnaiah has demonstrated a confidence bound analogue of condition II (see Corollary 3'(II) below) but has not pointed out the need for non-triviality of containment ([13], Theorem 6.4).

5. Simultaneous confidence statements. If the hypotheses of a family relate to parametric functions, simultaneous confidence statements may be made regarding the values of all those functions. The correspondence between such statements for certain functions and STP's for the hypotheses regarding particular values of these functions allows the extension of the foregoing properties of STP's to corresponding properties of these confidence statements. A number of additional definitions will be introduced to clarify this correspondence.

Consider a family \( \Phi = \{ \phi_i : i \in I \} \) of parametric functions, the range of \( \phi_i \) being denoted \( \Lambda_i \). A typical value of \( \phi_i(\theta) \) will be denoted \( \lambda_i \), it being understood that \( \lambda_i \in \Lambda_i \). Then

\[
\omega_i(\lambda_i) = \{ \theta : \phi_i(\theta) = \lambda_i \}
\]

is a hypothesis regarding the value of \( \phi_i(\theta) \), and the class \( \{ \omega_i(\lambda_i) : \lambda_i \in \Lambda_i \} \) of such hypotheses forms a disjoint partition of the parameter space \( \omega \).
For each value $\lambda_i$, let $Z_i(Y; \lambda_i)$ be a statistic whose distribution is the same for all $\theta \in \omega_i(\lambda_i)$ and is independent of the value of $\lambda_i$. Write $\tilde{Z}$ for the collection of all such statistics with $\lambda_i \in \Lambda_i, \ i \in I$. The collections of parametric functions $\phi_i$ and corresponding statistics $Z_i(Y; \lambda_i)$ will be denoted $[\Phi, \tilde{Z}]$ and referred to as an estimation family. For any given probability $\alpha_i$ let $\xi$ be such that

\[(5.2) \quad P_{\alpha_i,0}(Z_i(Y; \lambda_i) \leq \xi) = P_\theta(Z_i(Y; \phi_i(\theta)) \leq \xi) = 1 - \alpha_i,\]

this probability clearly being independent of the value of $\lambda_i$ or $\theta$. Then the subset of $\Lambda_i$

\[(5.3) \quad \Lambda_i(Y; \xi) = \{\lambda_i \mid Z_i(Y; \lambda_i) \leq \xi\}\]

is a confidence region for $\phi_i(\theta)$ with co-efficient $1 - \alpha_i$. The confidence statement

\[\phi_i(\theta) \in \Lambda_i(Y; \xi)\]

is equivalent to the statement $\theta \in \Theta_i(Y; \xi)$, with

\[(5.4) \quad \Theta_i(Y; \xi) = \{\theta \mid Z_i(Y; \phi_i(\theta)) \leq \xi\}\]

being a set in $\omega$. Either of these statements is understood in the sense of acceptance of hypothesis $\omega_i(\lambda_i)$ for any value $\lambda_i \in \Lambda_i(Y; \xi)$ by a test of level $\alpha_i$.

The family of statements

\[(5.5) \quad \{\phi_i(\theta) \in \Lambda_i(Y; \xi) \mid i \in I\} = \{\theta \in \Theta_i(Y; \xi) \mid i \in I\}\]

based on estimation family $[\Phi, \tilde{Z}]$ and critical value $\xi$ is referred to as a Simultaneous Confidence Statement (SCS) and denoted $[\Phi, \tilde{Z}, \xi]$. Clearly, all statements of $[\Phi, \tilde{Z}, \xi]$ are true if and only if $\theta \in \bigcap_i \Theta_i(Y; \xi)$, i.e., if the true parameters $\theta$ are in each one of the confidence regions $\Theta_i(Y; \xi), \ i \in I$. The probability that this event occurs, i.e.,

\[(5.6) \quad C = P_\theta(\phi_i(\theta) \in \Lambda_i(Y; \xi) \forall i \in I) = P_\theta(\theta \in \bigcap_i \Theta_i(Y; \xi)),\]

is referred to as the joint confidence co-efficient of $[\Phi, \tilde{Z}, \xi]$. By introducing (5.4) this may be written

\[(5.7) \quad C = P_\theta(\max \{Z_i(Y; \phi_i(\theta)) \mid i \in I\} \leq \xi)\].

It is important to distinguish a joint confidence region for several parametric functions from a simultaneous confidence statement. The former is a confidence region in the sense of (5.3), where $\phi_i$ is a vector valued parametric function and is equivalent to a single region in parameter space—(5.4). The latter consists of a family of such regions, each one relating to either a scalar or a vector $\Phi_i$. Note that Roy and Bose [22] have restricted consideration to the case where each $\phi_i$ was scalar valued and specifically to the case where each $\Lambda_i(Y; \xi)$ could be expressed as an interval on the real line.

**Example 5.1.** In the one-way ANOVA set-up a simultaneous confidence statement on all contrasts $c_j' y, c_j' \in V(C')$, in expectations is

\[\{c_j' y \in \Lambda_{c_j}(x, \xi') \forall c_j' \in V(C')\},\]
where $x$ indicates the observations, $\xi^p = (k - 1)F_{(k-1; n_2)_{1-\alpha}}$ and

$$
\Lambda_{ij}(x, \xi^p) = \{c_i' y \mid c_i' x - (\bar{\xi}^p s^2 c_i' N^{-1} c_i) \leq c_i' y \leq c_i' x + (\bar{\xi}^p s^2 c_i' N^{-1} c_i)^T\},
$$

with $N = \text{diag} \ (n_1, n_2, \ldots, n_k)$, is a confidence interval for contrast $c_i' y$. Furthermore, since $C_i' y$ generates all such contrasts

$$
\{C_i' y \mid \sum_{h=1}^k n_h (\bar{\xi}_h - \mu_h)^2 - (\sum_{h=1}^k n_h (\bar{\xi}_h - \mu_h))^2 / \sum_{h=1}^k n_h \leq \bar{\xi}^p s^2\}
$$

is a single joint confidence region for all contrasts in expectations. Since this region is the intersection of all the $\Lambda_{ij}(x, \xi^p)$ regions in the space of contrasts $C_i' y$ (and in that of the parameters $\mathbf{u}, \sigma^2$) the joint region and the simultaneous statement have, by (5.6), the same confidence co-efficient $1 - \alpha$.

This was first pointed out by Scheffé [25] (see also Example 3.3 for the related STP on null contrasts).

An estimation family is said to be related if the functions of $\Phi$ have a transitive relation $\geq$ which is neither symmetric nor reflexive, such that if $i < j$ then $\omega_i(\lambda_i)$ implies $\omega_j(\lambda_j)$ for a particular value $\lambda_j$, and $0 < i$ for all $i \in I - \{0\}$. For $i < j$, the relation $\geq$ defines a function $r_{ji}$ such that $\lambda_j = r_{ji}(\lambda_i)$. Since $\geq$ is transitive, $i < j$ and $j < g$ require $i < g$ so that if $\omega_i(\lambda_i) \in \omega_j(\lambda_j) \subseteq \omega_g(\lambda_g)$, $\lambda_g = r_{gi}(\lambda_i) = r_{gj}(\lambda_j) = r_{gi}(r_{ji}(\lambda_i))$, which imposes the relation $r_{gi} = r_{gj}(r_{ji})$ on these functions. From these definitions it follows that for $i < j$

$$
\omega_j(\lambda_j) = \bigcup_{\lambda_j = r_{ji}(\lambda_i) = \lambda_j \mid \omega_i(\lambda_i)}.
$$

(5.8)

The functions $\phi_{ij}$ for which there exists no $\phi_j \in \Phi$ such that $j < g$ will be indexed by $j \in I_{\text{min}}$. The adjective minimal will be applied to these functions and all other concepts depending on these functions, as, for instance, the confidence regions $\{\Lambda_j(\mathbf{y}; \xi) \mid j \in I_{\text{min}}\}$. Similarly, the adjective related will be applied to SCs based on related estimation families.

Example 5.2. In the subset ANOVA set-up one may be concerned with estimation of all contrasts in the expectations of each subset. For each subset $S_i$, $i \in I^s$, of $k_i$ expectations let $C_i$ be a $k_i \times (k_i - 1)$ matrix such that $C_i' y$ generates all contrasts of $S_i$. It is well known that whenever $S_j \subseteq S_i$ there exists a $B_{ji}$ such that $B_{ji}^T = B_{ji} C_i'$. Hence, if $C_i' y = \lambda_i$ this implies that $C_j' y = \lambda_j$ where $\lambda_j = B_{ji}^T \lambda_i$. Thus, the functions $\phi_{ij}(\mathbf{y}) = C_j' \mathbf{y}$ are defined for each $i \in I^s$ and a relation $\phi_{ij}(\mathbf{y}) = \hat{B}_{ji} \phi_{ij}(\mathbf{y})$ exists whenever $S_j \subseteq S_i$. An estimation family for these contrast sets will thus be related.

A simultaneous $1-\alpha$ confidence statement on all these linear sets of contrasts is

$$
\{C_i' y \in \Lambda_i(x, \xi^p) \mid i \in I^s\}
$$

where, with the same definitions as before,

$$
\Lambda_i(x, \xi^p) = \{C_i' y \mid (\bar{x} - y)' C_i (C_i' N^{-1} C_i)^{-1} C_i' (\bar{x} - y) \leq \xi^p s^2\}
$$

$$
= \{C_i' y \mid \sum_{h \in S_i} n_h (\bar{x}_h - \mu_h)^2 - (\sum_{h \in S_i} n_h (\bar{x}_h - \mu_h))^2 / \sum_{h \in S_i} n_h \leq \xi^p s^2\}
$$

is the confidence region for $C_i' y$, using the augmented $F$-ratio.
The corresponding STP for hypotheses of null contrast sets was given in Examples 3.1 and 3.4.

Example 5.3. In the same set-up as that of Example 5.2 Roy and Gnanadesikan [23] have proposed putting confidence bounds on non-centrality parameter
\[
\phi_i(\mathbf{y}) = \{\mathbf{y}'C_i[N^{-1}C_i]'C_i'y\}^{1/2}
\]
for each \(i \in I^S\). It is readily checked that if \(S_i \subseteq S_j\), then for any \(\lambda_i, \phi_i(\mathbf{y}) = \lambda_i\) does not generally imply a particular single value \(\lambda_j\) for \(\phi_j(\mathbf{y})\), (except that if \(\lambda_i = 0\) then also \(\lambda_j = 0\)). Hence these non-centrality parameter functions do not allow of a related estimation family.

Defining statistics
\[
Y_i = \mathbf{x}'C_i[N^{-1}C_i]'C_i'y
\]
for each \(i \in I^S\), the Roy-Gnanadesikan confidence statement is \(\{\phi_i(\mathbf{y}) \in \Lambda_i^{R\delta}\} \cap \{\mathbf{x}, \xi^P \mid \xi \in I^S\}\), with the confidence regions
\[
\Lambda_i^{R\delta}(\mathbf{x}, \xi^P) = \{\lambda_i \mid Y_i^\lambda - (\xi^P s^2)^{1/2} \leq \lambda_i \leq Y_i^\lambda + (\xi^P s^2)^{1/2}\}.
\]
The sense in which this SCS is non-related is illustrated in Example 5.5, below.

For a related estimation family \([\Phi, \mathcal{Z}]\) consider, for any \(\lambda_0 \in \Lambda_0\), the family of hypotheses
\[
(5.9) \quad \Omega_{\lambda_0} = \{\omega_i(\lambda_i) \mid \lambda_i = r_{i0}(\lambda_0)\},
\]
and the family of statistics
\[
(5.10) \quad Z_{\lambda_0} = \{Z_i(Y; \lambda_i) \mid \lambda_i = r_{i0}(\lambda_0)\}.
\]
It is easily checked that \(\mathcal{J}\) is an implication relation for \(\Omega_{\lambda_0}\) and that \(\{\Omega_{\lambda_0}, Z_{\lambda_0}\}\) satisfies the requirements of a testing family (Section 2, above). Hence \(\{\Omega_{\lambda_0}, Z_{\lambda_0}, \xi\}\) is an STP of level
\[
(5.11) \quad \alpha = P_{\omega_0(\lambda_0)}(Y(Y; \lambda_0) > \xi).
\]
This STP accepts \(\omega_i(\lambda_i)\), where \(\lambda_i = r_{i0}(\lambda_0)\), if \(Z_i(Y; \lambda_i) \leq \xi\). Thus, acceptance of \(\omega_i(\lambda_i)\) by STP \(\Omega_{\lambda_0}, Z_{\lambda_0}, \xi\), where \(\lambda_i = r_{i0}(\lambda_0)\), is determined by the same event as inclusion of the parametric value \(\lambda_i\) in SCS \([\Phi, \mathcal{Z}, \xi]\), i.e., \(\lambda_i \in \Lambda_i(Y; \xi)\) as in (5.3). This brings out the correspondence between the STP’s \(\Omega_{\lambda_0}, Z_{\lambda_0}, \xi\) for all \(\lambda_0 \in \Lambda_0\) and the related SCS \([\Phi, \mathcal{Z}, \xi]\). The experimentwise error rate of \(\Omega_{\lambda_0}, Z_{\lambda_0}, \xi\) is, provided \(Z_0(Y; \lambda_0) = \max \{Z_i(Y; \lambda_i) \mid \lambda_i = r_{i0}(\lambda_0)\}\) (see (5.16) below),
\[
\alpha = P_{\omega_0(\lambda_0)}(\max \{Z_i(Y; \lambda_i) \mid \lambda_i = r_{i0}(\lambda_0)\} > \xi)
= P_{\theta}(\max \{Z_i(Y; \phi_i(\theta)) \mid i \in I\} > \xi)
\]
which is seen to be \(1 - C\), for joint confidence co-efficient \(C\).

Implication relations have been defined for a related estimation family;
hence the terms coherence and consonance may also be defined for a related SCS. For non-related SCS's these terms have no clear meaning. Coherence was defined in Section 3 as the requirement that if \( \omega_i(\lambda_i) \subset \omega_j(\lambda_j) \), acceptance of \( \omega_i(\lambda_i) \) would entail acceptance of \( \omega_j(\lambda_j) \). Consonance was defined as the requirement that \( \omega_i(\lambda_i) \) be accepted if all \( \omega_j(\lambda_j) \) were accepted for which \( \omega_i(\lambda_i) \subset \omega_j(\lambda_j) \). Correspondingly, related SCS \( [\Phi, \tilde{Z}, \tilde{\xi}] \) is said to be coherent if, when \( \lambda_j = r_{ji}(\lambda_i) \) then \( \lambda_i \in \Lambda_i(y; \tilde{\xi}) \) ensures \( \lambda_j \in \Lambda_j(y; \tilde{\xi}) \). It is said to be consonant if \( \lambda_j \in \Lambda_j(y; \tilde{\xi}) \forall \lambda_j = r_{ji}(\lambda_i) \) ensures \( \lambda_i \in \Lambda_i(y; \tilde{\xi}) \). Thus \( [\Phi, \tilde{Z}, \tilde{\xi}] \) is coherent and consonant when the statements \( \lambda_i \in \Lambda_i(y; \tilde{\xi}) \) and \( \lambda_j \in \Lambda_j(y; \tilde{\xi}) \) for all \( j \) such that \( \lambda_j = r_{ji}(\lambda_i) \) must occur together for all \( i \in I \).

In terms of the \( \theta \) set formulation the coherence requirement is that for all \( i \in I \)

\[
(5.12) \quad \Theta_i(y; \tilde{\xi}) \subset \bigcap_{j|j < i} \Theta_j(y; \tilde{\xi})
\]

and the consonance requirement is that for all \( i \in I \)

\[
(5.13) \quad \Theta_i(y; \tilde{\xi}) \supset \bigcap_{j|j < i} \Theta_j(y; \tilde{\xi}).
\]

Thus \( [\Phi, \tilde{Z}, \tilde{\xi}] \) is coherent and consonant if

\[
(5.14a) \quad \Theta_i(y; \tilde{\xi}) = \bigcap_{j|j < i} \Theta_j(y; \tilde{\xi})
\]

for every \( i \in I \).

In view of the analogous definitions of coherence and consonance for STP's and SCS's, it is immediately seen that a SCS is coherent and/or consonant, if and only if every one of the corresponding STP's is coherent and/or consonant, respectively.

Finally, to complete the analogy between \( [\Phi, \tilde{Z}, \tilde{\xi}] \) and \( \{\Omega_\lambda, Z_\lambda, \xi\} \) for all \( \lambda_0 \in \Lambda_0 \), the estimation family \( [\Phi, \tilde{Z}] \) will be called monotone if

\[
(5.14b) \quad Z_i(y; \lambda_i) \geq Z_j(y; \lambda_j) \quad \text{a.e.}
\]

whenever \( \lambda_j = r_{ji}(\lambda_i) \), and strictly monotone if

\[
(5.15) \quad Z_i(y; \lambda_i) = \max \{Z_j(y; \lambda_j) \mid \lambda_j = r_{ji}(\lambda_i)\} \quad \text{a.e.}
\]

for all \( i \in I \). In particular, for any monotone \( [\Phi, \tilde{Z}] \)

\[
(5.16) \quad Z_0(Y; \lambda_0) = \max \{Z_i(Y; \lambda_i) \mid \lambda_i = r_{i0}(\lambda_0)\}.
\]

These definitions correspond, for any \( \{\Omega_\lambda, Z_\lambda\} \) to (2.1), (2.2) and (2.4), so that (strict) monotonicity of \( [\Phi, \tilde{Z}] \) corresponds to the same property of \( \{\Omega_\lambda, Z_\lambda\} \) for each \( \lambda_0 \in \Lambda_0 \). Also, estimation family \( [\Phi, \tilde{Z}] \) will be said to be UI related if testing family \( \{\Omega_\lambda, Z_\lambda\} \) is UI related for each \( \lambda_0 \).

With these analogous definitions and properties of STP's and SCS's it is clear that the theorems and corollaries that have been proved for STP's will apply, mutatis mutandis, to SCS's. They will be stated below without further proof; the number of each statement being the same as for STP's except for a prime.
Theorem 1'. All SCS's based on [Φ, Z] are coherent if and only if [Φ, Z] is monotone.

Corollary 1'. If Z_i(Y; λ_i) is a LR statistic for ω_i(λ_i) for all Z_i(Y; λ_i) ε Z, then all SCS's based on [Φ, Z] are coherent.

Theorem 2'. The joint confidence co-efficient of coherent [Φ, Z, ζ] is C = 1 - α, for α of (5.11). The confidence co-efficient of any φ_i is at least 1 - α.

Theorem 3'. All SCS's based on [Φ, Z] are coherent and consonant if and only if [Φ, Z] is strictly monotone.

Corollary 2'. All SCS's based on [Φ, Z] are coherent and consonant if, and only if, [Φ, Z] is UI related.

Example 5.4. To show the monotonicity of the estimation family of Example 5.2 and the coherence of the resulting SCS's one may argue as follows. It follows from a well known theorem (applying 1f.1.1 of [19] with c' = x'C_j'), that

\[ \max_{c' \in V(c, r)} [c'(\bar{x} - \mu_0)]^2 / s^2 c'N^{-1}c = (C_j' \bar{x} - C_j'\mu_0)(C_j'N^{-1}C_j)^{-1}(C_j' \bar{x} - C_j'\mu_0) / s, \]

where the right hand side is Z_i(Y; C_j'\mu_0), the statistic for ω_i in Example 5.2. Now, hypotheses ω_i(λ_i) : C_j'\mu = λ_i have implication relations ω_i(λ_i) ⊆ ω_j(λ_j) for all S_i ⊃ S_j, provided these λ_i, i ε I^8 are such that there exists a ω_j for which C_j'\mu_0 = λ_j,∀i ε I^8. Further, S_i ⊃ S_j is equivalent to V(C_j') ⊃ V(C_j') so that

\[ \max_{c' \in V(c, r)} [c'(\bar{x} - \mu_0)]^2 / s^2 c'N^{-1}c \geq \max_{c' \in V(c, r)} [c'(\bar{x} - \mu_0)]^2 / s^2 c'N^{-1}c, \]

i.e., Z_i(Y; λ_i) ≥ Z_j(Y; λ_j), which proves monotonicity.

Coherence of the SCS of Example 5.2 follows by a similar argument. Let S_i ⊃ S_j, let λ_i be a value of C_j'\mu and write \mu_0 for any value such that λ_i = C_j'\mu_0. Now λ_i ε \Lambda_i(\mu, ζ) is equivalent to saying that Z_i(\mu; C_j'\mu_0) ≤ ζ. Hence, by the inequality above, also Z_i(\mu; C_j'\mu_0) ≤ ζ, which is equivalent to saying that λ_i ε \Lambda_j(\mu, ζ), where λ_j = C_j'\mu_0, the value corresponding to λ_i = C_j'\mu_0. Inclusion of λ_i in the SCS implies inclusion of λ_j, establishing coherence.

Example 5.5. In the Roy-Gnanadesikan set-up of Example 5.3 one notes, arguing as in Example 5.4, that if S_i ⊃ S_j then for any ω, \phi_i(ω) ≥ \phi_j(ω) and Y_i ≥ Y_j. Now, for any given ω the random event

\[ Y_j^t + (\zeta^2 s^2)^{1/4} ≥ \phi_i(ω) ≥ \phi_j(ω) ≥ Y_j^t + (\zeta^2 s^2)^{1/4}, \]

may occur. In that case \phi_i(ω) ε \Lambda_i^{R,0}(\mu, ζ) but \phi_j(ω) ε \Lambda_j(\mu, ζ). Thus for a given set of expectations ω, the non-centrality parameter \phi_i(ω) may be included in the Roy-Gnanadesikan SCS whereas the other non-centrality parameter \phi_j(ω) is not. Note also that in the even that

\[ \phi_i(ω) ≥ Y_i^t + (\zeta^2 s^2)^{1/4} ≥ Y_j^t + (\zeta^2 s^2)^{1/4} ≥ \phi_j(ω) \]

the converse would occur, \phi_j(ω) being included whereas \phi_i(ω) was excluded. Evidently there is no coherence in the inclusions within this SCS.

Example 5.6. An alternative approach may be taken in the subset ANOVA
set-up if sample sizes are equal. For any subset $S_i$ of $k_i$ expectations $\mu_{i_1}, \mu_{i_2}, \ldots, \mu_{i_{k_i}}$ one may define $\phi_i^S(y)$ as the set of $(y^i)$ differences $\mu_{i_a} - \mu_{i_b}, i_a \neq i_b$. For any values $\lambda_i$ of the differences in set $\phi_i^S(y)$ one may define statistic

$$Z_i^R(x, \lambda_i) = n_i^4 \max \{[\bar{x}_{i_a} - \bar{x}_{i_b} - (\mu_{i_a} - \mu_{i_b})]/s_i | \mu_{i_a}, \mu_{i_b} \in S_i\}.$$ 

The distribution of $Z_i^R(x, \lambda_i)$ under $\omega_i(\lambda_i) = \{y_i | \phi_i^S(y) = \lambda_i\}$ is that of a studentized range of $k_i$ means and $n_e$ d.f. irrespective of the value of $\lambda_i$. Hence $[\Phi^S, \bar{Z}^R]$ is an estimation family.

Defining, for any $i, j$ such that $S_i \supseteq S_j$, $r_{ji}$ as the function picking out the subset $\phi_j^S(y)$ from the set $\phi_i^S(y)$ of differences, $\phi_j^S(y) = r_{ji}(\phi_i^S(y))$ whenever $S_i \supseteq S_j$. Now, if $\phi_i^S(y) = \lambda_i$, clearly $\phi_j^S(y) = r_{ji}(\lambda_i)$, so that $\omega(\lambda_i) \subset \omega_j(\lambda_i)$ if $\lambda_j = r_{ji}(\lambda_i)$. Hence the estimation family $[\Phi^S, \bar{Z}^R]$ is related, and so is the SCS.

The resulting SCS of confidence co-efficient $1 - \alpha$ is $\{\phi_i^S(y) \in \Lambda_i^R(x, \xi^R) | i \in I^g\}$ where $\xi^R$ is the upper $\alpha$ point of the $k$ mean studentized range distribution with $n_e$ d.f. and

$$\Lambda_i^R(x, \xi^R) = \{\lambda_i | Z_i^R(x, \lambda_i) \leq \xi^R\}.$$ 

Note, in particular, that for $S_j$ consisting of just two expectations $\phi_j^S(y) = \mu_{i_a} - \mu_{i_b}$, and the confidence region becomes the interval

$$\Lambda_j^R(x, \xi^R) = \{\lambda_j | \bar{x}_{i_a} - \bar{x}_{i_b} - \xi^R s_i/n_i^4 \leq \mu_{i_a} - \mu_{i_b} \leq \bar{x}_{i_a} - \bar{x}_{i_b} + \xi^R s_i/n_i^4\}$$

on the value of the difference $\lambda_j = \mu_{i_a} - \mu_{i_b}$. This type of SCS is due to Tukey [28], being referred to as the method of allowances. The corresponding STP for null differences was discussed in Example 3.7.

Since $\phi_i^S(y) \supseteq \phi_j^S(y)$ for $S_i \supseteq S_j$ it is clear that $Z_i^R(x, \lambda_i) \geq Z_j^R(x, \lambda_j)$ provided $\lambda_j = r_{ji}(\lambda_i)$. Moreover, it is readily apparent that

$$Z_i^R(x, \lambda_i) = \max \{Z_j^R(x, \lambda_j) | \lambda_j = r_{ji}(\lambda_i), S_j \subset S_i\}$$

so that $[\Phi^S, \bar{Z}^R]$ is strictly monotone.

Coherence and consonance of SCS $[\Phi^S, \bar{Z}^R, \xi^R]$, which follows from Theorem 3', may be checked as follows. From the above relation between $Z_i^R(x, \lambda_i)$ for $S_i$ and the statistics $Z_j^R(x, \lambda_i)$ for all pairs of expectations $S_j(\subset S_i)$, it follows that $(Z_i^R(x, \lambda_i) \leq \xi^R)$ is equivalent to $(Z_j^R(x, \lambda_j) \leq \xi^R)$, where $\lambda_j = r_{ji}(\lambda_i), \forall$ pairs $S_j(\subset S_i)$. Thus the statement $\lambda_i \in \Lambda_i^R(x, \xi^R)$ is equivalent to the statement $\lambda_j \in \Lambda_j^R(x, \xi^R)$ where $\lambda_j = r_{ji}(\lambda_i)$, for all pairs $S_j(\subset S_i)$, and this establishes coherence and consonance.

Resolution may be defined and compared for SCS's analogously to the way it was for STP's in Section 4, above. However, a comparison of SCS's $[\Phi^S, \bar{Z}^R, \xi^R]$ and $[\Phi^S, \bar{Z}^R, \xi^R]$ has meaning only if the hypotheses on the functions $\Phi$ are the same as those on some of the functions of $\Phi^S$. Thus, for example, for each value $\lambda_0$ of $\phi_0 \in \Phi$ there must exist another value $\lambda_0^*$, say, of $\phi^* \in \Phi^*$ such that $\omega_0(\lambda_0) = \omega_0^*(\lambda_0^*)$. If that is so and if $I_{min} \subseteq I^R$ then the conditions

$$(5.17a) \quad P_{\omega_0(\lambda_0)}(Z_0(Y; \lambda_0) \leq \xi^R) = P_{\omega_0^*(\lambda_0^*)}(Z_0^*(Y; \lambda_0^*) \leq \xi^*)$$
equally for all \( \lambda_0 \in \Lambda_0 \), and

\[
(5.17b) \quad \Lambda_j(y; \xi) \subseteq \Lambda_j^*(y; \xi^*) \quad \text{a.e.,}
\]

define \([\Phi, \tilde{Z}, \xi]\) to be no less resolvent than \([\Phi^*, \tilde{Z}^*, \xi^*]\). If the containment in (5.17b) is proper the former SCS is said to be strictly more resolvent than the latter. Thus, if two SCS's which can be compared have the same joint confidence co-efficient, but the minimal confidence regions of the one are contained in the corresponding regions of the other, the former is said to be more resolvent.

Again, let estimation families \([\Phi, \tilde{Z}, \xi]\) and \([\Phi^*, \tilde{Z}^*, \xi^*]\) have related hypotheses, and for each \( \lambda_0 \in \Lambda_0 \) and the corresponding \( \lambda_0^* \) such that \( \omega_0(\lambda_0) = \omega_0^*(\lambda_0^*) \), let \( \Omega_{\lambda_0^*} \subseteq \Omega_{\lambda_0} \), also hold. Then \([\Phi, \tilde{Z}, \xi]\) will be said to be no less parsimonious than \([\Phi^*, \tilde{Z}^*, \xi^*]\) if

\[
(5.18a) \quad \text{for all } j \in I_{\text{min}} \Lambda_j(y; \xi) = \Lambda_j^*(y; \xi^*) \quad \text{a.e.,}
\]
\[
(5.18b) \quad \text{for all } i \in I \Lambda_i(y; \xi) \subseteq \Lambda_i^*(y; \xi^*) \quad \text{a.e.,}
\]
\[
(5.18c) \quad \text{and for } \omega_0(\lambda_0) = \omega_0^*(\lambda_0^*) \quad C \geq C^*.
\]

Again, if the latter inequality is strict the former SCS is said to be strictly more parsimonious than the latter. In other words, for equal minimal confidence regions the more parsimonious SCS has all non-minimal confidence regions contained in the corresponding regions of the less parsimonious SCS, and its joint confidence co-efficient is larger. Parsimony must here be understood in the sense of increasing joint confidence, and that is equivalent to reducing the experiment-wise error rate.

Further, estimation families with related hypotheses may allow narrowness comparisons. Thus \([\Phi, \tilde{Z}]\) is said to be narrower than \([\Phi^*, \tilde{Z}^*]\) if, for each \( \lambda_0 \) and the corresponding \( \lambda_0^* \) (i.e., \( \omega_0(\lambda_0) = \omega_0^*(\lambda_0^*) \)), \( \Omega_{\lambda_0}, \Omega_{\lambda_0}^* \) is narrower than \( \Omega_{\lambda_0^*}, \Omega_{\lambda_0}^* \). The requirements for this are that for each \( \lambda_0 \) there exist a function \( g_{\lambda_0} \) satisfying (4.3a, b, c, d) for the corresponding hypotheses. This function may or may not be the same for all \( \lambda_0 \in \Lambda_0 \). Thus, narrowness of estimation families is defined simply as narrowness of the testing families which they generate.

Again, the following theorem and corollaries may be stated in view of these further analogies between properties of STP's and SCS's.

**Theorem 4'**. If \([\Phi, \tilde{Z}]\) is narrower than \([\Phi^*, \tilde{Z}^*]\), then for every \( \xi^* \) there exists \( \xi \) such that \([\Phi, \tilde{Z}, \xi]\) is no less parsimonious than \([\Phi^*, \tilde{Z}^*, \xi^*]\). Moreover, the former SCS is strictly more parsimonious than the latter for some values of \( \xi^* \), i.e., those for which

\[
P_{\omega_0(\lambda_0)}(Z^*_0(Y; \lambda_0) > \xi^* \geq g_{\lambda_0}(Z_0(Y; \lambda_0))) > 0,
\]

where \( g_{\lambda_0} \) satisfies (4.3a, b, c, d) for all \( \lambda_0 \in \Lambda_0 \).

**Theorem 5'**. If \([\Phi, \tilde{Z}]\) is narrower than \([\Phi^*, \tilde{Z}^*]\) and if \([\Phi, \tilde{Z}, \xi]\) and \([\Phi^*, \tilde{Z}^*, \xi^*]\) have the same co-efficient, then the former SCS is no less resolvent than the latter. Moreover, it is strictly more resolvent for some values of \( \xi \), i.e., those for which

\[
P_{\omega_0(\lambda_0)}(Z^*_0(Y; \lambda_0) > g_{\lambda_0}(\xi) \geq g_{\lambda_0}(Z_0(Y; \lambda_0))) > 0
\]

where \( g_{\lambda_0} \) satisfies (4.3a, b, c, d) for all \( \lambda_0 \in \Lambda_0 \).
Clearly, narrower estimation families have the advantage of providing more resolvent SCS's with the same joint confidence co-efficient and more parsimonious SCS's for identical minimal confidence regions.

**Example 5.7.** In a subset ANOVA set-up with equal sample sizes, compare the estimation families \([\Phi^*, \widehat{Z}^*]\) of Example 5.6 and \([\Phi^*, \bar{Z}^*]\) of Example 5.2. The relation \(Z' = g(Z^*) = (Z^*)^2/2\) holds for all \(Z_j' (x, \lambda_j)\) and \(Z_j^* (x, \lambda_j)\) if \(S_j\) is a pair. As in Example 4.4, above, so also here \(Z_j' (x, \lambda_i) > [Z_j^* (x, \lambda_i)]^2/2\) if \(S_i\) contains three or more expectations and these inequalities are strict a.e. Hence \([\Phi^*, \bar{Z}^*]\) is narrower than \([\Phi^*, \bar{Z}^*]\) and it follows from Theorem 4' and 5' that for all subsets the SCS's due to Tukey are more parsimonious or more resolvent than the corresponding subset SCS's based on Scheffé's approach.

**Example 5.8.** The comparison in Example 5.7 between Tukey and Scheffé confidence statements for subsets of expectations also follows, by virtue of Corollary 3', condition (I), from the UI and LR relations of the respective estimation families, and the fact that \(Z_j' (x, \lambda_j) = [Z_j^* (x, \lambda_j)]^2/2\) for all pairs \(S_j\).

**Corollary 3'.** Let \([\Phi, \bar{Z}]\) and \([\Phi^*, \bar{Z}^*]\) be monotone estimation families such that \(\Phi \subseteq \Phi^*, \Phi_{\min} \subseteq \Phi_{\min}^*\) and \(\omega_0 (\lambda_0) = \omega_0^* (\lambda_0^*)\) for all \(\lambda_0 \in \Lambda_0\) and let there exist for each \(\lambda_0 \in \Lambda_0\) a function \(g_{\lambda_0}\) satisfying (4.3a, b) for the corresponding hypotheses. Then each of the following are sufficient conditions that \([\Phi, \bar{Z}]\) provide more parsimonious and more resolvent SCS's than \([\Phi^*, \bar{Z}^*]\).

1. If \([\Phi, \bar{Z}]\) is UI related and \(g_{\lambda_0}\) also satisfy (4.3e) or (4.3e') for each \(\lambda_0\),
2. If both \([\Phi, \bar{Z}]\) and \([\Phi^*, \bar{Z}^*]\) are UI related and

\[
 P_{\omega_0, \lambda_0} (\max \{Z_j^* (Y; r_j, \lambda_0) \mid j \in I_{\min} \} > \max \{Z_j^* (Y; r_j, \lambda_0) \mid j \in I_{\min} \} > 0
\]

for all \(\lambda_0 \in \Lambda_0\).

This corollary covers some results of Krishnaiah for ANOVA and MANOVA set-ups [13]. He uses UI related families throughout and his Theorem 6.2, 6.3 and 6.4 establish that for an identity function \(g\) and finite \(I_{\min} \subseteq I_{\min}^*\), the minimal confidence regions of \([\Phi, Z, \xi]\) will be properly contained in those of \([\Phi^*, \bar{Z}^*, \xi^*]\) where both SCS's have the same co-efficient. Apart from cases where (5.19) might be zero his conditions are seen to correspond to (II) of Corollary 3', and his containment conclusion corresponds, again apart from possible sets of probability zero, to the greater resolution conclusion of Corollary 3'. He does not explicitly discuss anything corresponding to parsimony.

The following are two examples of Krishnaiah's results:

**Example 5.9.** Dunnott's simultaneous \(1 - \alpha\) confidence bounds on comparisons of the expectations under \((k - l)\) treatments with that under a control [5] are more resolvent than Tukey's simultaneous \(1 - \alpha\) confidence bounds on all pairwise comparisons of these \(k\) expectations [27]. (Krishnaiah, [13], Corollary 6.3).

The corresponding comparison of STP's was given in Example 4.13.

**Example 5.10.** Simultaneous SANOVA \(1 - \alpha\) confidence bounds on a finite
subset of contrasts in \( k \) expectations are more resolvent than Scheffé-type simultaneous 1 \( - \alpha \) confidence bounds on all contrasts (Krishnaiah [13], Theorem 6.2).

6. Other methods of multiple comparisons. The previous discussion relates almost entirely to STP’s and SCS’s which are based on monotone families, and thus provide coherent decisions. Some examples have been mentioned of procedures for non-monotone families in which the decisions are not always coherent and the properties in the theorems do not apply. In particular, we may mention the use of the Kruskal-Wallis statistic for nonparametric ANOVA (Example 3.2) proposed by Nemenyi [17] (see also the Miller’s discussion [16], Section 4.6). One would presume that incoherences, though possible with this technique, are pretty rare, especially with large samples. In that case such a technique might be said to be asymptotically coherent. This way warrant further investigation.

Other techniques have a step-wise procedure built in which ensures coherence but destroys simultaneity. Thus, Newman [18] and Keuls [10], in the equal sample size ANOVA, decide on each subset of means according to whether the studentized range for that subset or any other set containing it is \( \alpha \) significant in the distribution of its range. Duncan [3] proceeds in a more intricate way, choosing a different percentage point for each one of the ranges. He also extended the method to \( F \)-statistics [4].

In these methods the decision on a particular \( \omega_i \) depends not only on the statistic \( Z_i \), but on all \( Z_j \) such that \( j < i \). Clearly, this does not provide a testing family in the sense of Section 2 and the STP results do not apply. These methods have the disadvantage of making the decision on \( \omega_i \) depend in part on statistics \( Z_j \), whose distribution depends not only on \( \omega_i \) but also on \( \omega_j - \omega_i \) which is irrelevant to \( \omega_i \). On the other hand, these methods offer better resolution—in the sense of Section 4—than STP’s and are therefore preferred by some workers. (For a detailed discussion see [6], Section 9.)

Acknowledgment. I am indebted to my colleagues in Jerusalem and Chapel Hill with whom I have discussed these techniques. I am particularly happy to acknowledge the help of J. Putter who critically read an early version of the paper and that of W. J. Hall who made many useful suggestions in the last revisions of this paper.

REFERENCES


