AD

Technical Memorandum 8-90

A NEW DIAGNOSTICS APPROACH FOR REPEATED MEASURES DESIGNS

Jock O. Grynovicki

June 1990
AMCMS Code 612716.H700011

U. S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland

®VAX is a registered trademark of Digital Equipment Corporation.

®SAS is a registered trademark of SAS Instituted Incorporated.

™SPSS^X is a a registered trademark of SPSS Incorporated.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Memorandum 8-90	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Human Engineering Laboratory	6b. OFFICE SYMBOL (If applicable) SLCHE
7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State, and ZIP Code) Aberdeen Proving Ground, MD 21005-5001	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
6.27.16	1L162716AH70		

11. TITLE (Include Security Classification)
A New Diagnostics Approach For Repeated Measures Designs

12. PERSONAL AUTHOR(S)
Grynovicki, Jock O.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM ____ TO ____	1990, June	38

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
01	03	01
12	03 (see reverse)	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
repeated measures design; variance component estimates; diagnostics; within-subject interactions; compound symmetry; sphericity (see reverse)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The traditional univariate analysis of the repeated measures design is obtained by treating subjects and their associated interactions as random effects. This analysis requires that certain variances and covariances of the dependent variable at various combinations of within-subject factors are equal. Instability of the variance and covariance components may mask significant effects and may compel the researcher to use a less powerful multivariate technique, provided sufficient subjects are available. The number of subjects required in the multivariate approach is a function of the number of dependent variables and levels of the within-subject factors. Thus, the researcher may be forced to use a conservative degree-of-freedom adjustment instead, which may mask significant results if subject availability is limited.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Technical Reports Office	22b. TELEPHONE (Include Area Code) (301)278-4478
22c. OFFICE SYMBOL SLCHE-SS-TSB	

DD Form 1473, JUN 86 — Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

17. (continued)

23 02

18. (continued)

"ave" estimators
subject interactions
residual analysis

19. (continued)

This report illustrates the use of a recently developed class of unbiased variance component estimators and their associated diagnostics for examining the data and the model assumptions. The researcher, with the use of these diagnostics, will be able to identify the source of the sphericity violation. Consequently, by modifying the model to account for unexplained sources of variability or by removing outliers that may cause sphericity violations, a more powerful univariate analysis can be performed.

AMCMS Code 612716.H700011 Technical Memorandum 8-90

A NEW DIAGNOSTICS APPROACH FOR REPEATED MEASURES DESIGNS

Jock O. Grynovicki

June 1990

APPROVED: JOHN D. WEISZ
Director
Human Engineering Laboratory

U.S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland 21005-5001

CONTENTS

A NEW DIAGNOSTICS APPROACH FOR REPEATED MEASURES DESIGNS

INTRODUCTION

The traditional univariate analysis of the repeated measures design is obtained by treating subjects and their associated interactions as random effects. This analysis requires the variance-covariance matrix of the response vector to have a specific form. Suppose that the treatment factors are arbitrarily divided into two groups, A and B. It is commonly assumed that the covariances between responses on the same subject and at the same levels of the factors in group A and at different levels of the factors in group B are the same, regardless of which levels are chosen for group A factors or which pairs of distinct levels of group B factors are chosen. For example, in a three-way completely random factorial model, Cov $(\bar{Y}ijkm, \bar{Y}ij^{*}k^{*}m^{*})$ is assumed constant (ϕ_1) for all i and each combination of $j \neq j^{*}$, $k \neq k^{*}$ (regardless of m and m^{*}). Instability of the variance and covariance components may mask significant effects or may compel the researcher to use a less powerful multivariate technique.

PURPOSE

This report illustrates the use of a recently developed class of unbiased variance component estimators and its associated diagnostics for examining the data and the model assumptions. The Repeated Measures Methodology section provides a review of the traditional methods of analyzing a repeated measures design. In the General Variance Component Estimates and Diagnostics Methodology section, to make the development and extension of the diagnostics to repeated measures comprehensible, appropriate notation and terminology is introduced. Finally, a comprehensive example is given for the case of a three-way design with two factors repeated.

REPEATED MEASURES METHODOLOGY

Repeated measures designs are some of the most frequently used classes of designs in psychology, physics, psychophysiology, and medicine. These designs offer a reduction in the error variance because of the removal of an individual's variability. They are efficient and require fewer subjects to achieve the same power of the F-test as completely random or block designs.

This class of designs, sometimes referred to as within-subject designs, obtains its name because one or more factors of the design are manipulated in such a way that each subject receives all levels of the within-subject factor. An advantage of this approach is that subjects act as their own control in their responsiveness to the various experimental treatments. On the other hand, this type of design introduces intercorrelations among the means upon which the test of within-subject main effects and interactions is based.

Because of this intercorrelation between the repeated treatments, three separate approaches have been proposed in the literature. The first, the univariate analysis of the repeated measures design, is obtained by treating subjects as a random effect. Second, the linear model employed is called a mixed effects model, and the resulting analysis is a mixed model analysis of the repeated measures design. Finally, the standard mixed model assumes certain variances and covariances of responses are invariant across the experiment.

In a three-factor factorial model with Factors 1 and 3 fixed and subjects (or Factor 2) random, a standard assumption is that the covariance, θ_{12}, of responses at the same level of Factor 1 and on the same subject (i.e., level of Factor 2) but at different levels of Factor 3, is invariant across all subjects, all levels of Factor 1 and all combinations of distinct levels of Factor 3. For example, for each level i of Factor 1 and subject j, the covariance of the a_3 x 1 vector $(\bar{Y}ijk.)_k' = (\bar{Y}ij1., \bar{Y}ij2., \ldots, \bar{Y}ija_3.)$ has the form $(\theta_{123} - \theta_{12})\, I_{a_3} + \theta_{13} J_{a_3} J_{a_3}$, in which I_{a_3} is an a_3 x a_3 identity matrix, and J_{a_3} is an a_3 x 1 vector of ones. The same covariance structure applies regardless of the value of i and j. Analogous statements apply to the a_1 by 1 vector $(\bar{Y}ijk.)_i$ for fixed j and k. This assumption is referred to in the literature as compound symmetry. More generally, for any design, if θ_t is the covariance between observations at the same levels of factors indexed by t and at different levels of the other factors, standard mixed models assume that θ_t is invariant across all levels of the factors indexed by t and across all combinations of distinct levels of the other factors.

A more general case in which the mean square ratio of a two-factor model has an exact F distribution is described in Huynh and Feldt (1970). This condition, referred to as sphericity, requires that $C'\Sigma C = \sigma^2 I$, in which C is a (k-1) x k orthonormal contrast matrix, I is the identity matrix of rank (k-1), and Σ is the variance-covariance matrix. An alternate representation of this necessary and sufficient condition is that

$$\Sigma = \lambda I + \alpha J' + J\alpha', \text{ in which } \alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_{a_3}).$$

Specifically, this can be written $\Sigma = (\sigma_{ij}) = (\sigma_{ii} + \sigma_{jj} - \sigma_{..} + \lambda(\delta_{ij} - 1/k))$,

in which $\lambda > 0$, $\delta_{ij} = 1$ if i = j and 0 otherwise.

It should be noted that compound symmetry is a special case of sphericity. It has been stated that, although it is not necessary, the absence of compound symmetry indicates that sphericity is unlikely (O'Brien & Kaiser, 1985). Several papers have appeared that state the assumption of sphericity is frequently violated. The consequence of such a violation is positive bias, meaning that the likelihood of a Type I error exceeds the nominal level, alpha (α).

To test the sphericity assumption, or equivalently, the Huynh and Feldt Type H pattern, one uses the Mauchly (1940) test statistic. A weighted function of this statistic has a chi-square distribution with (1/2)p(p-1)-1 degrees of freedom. Here, p is the number of treatment levels.

To compensate for nonsphericity, a degree-of-freedom adjustment (ε), initially proposed for use by Greenhouse and Geisser (1959), is used to adjust the ratio of the numerator and denominator degrees of freedom. Huynh and Feldt (1970) have shown this adjustment to be too conservative.

Values of ε range from 1, indicating sphericity to $1/(k-1)$, indicating maximum nonsphericity. Since the population Σ is rarely known, one approach proposed by Greenhouse and Geisser is to use the lower limit $1/(k-1)$. Huynh and Feldt have shown this adjustment to be conservative.

In the multivariate method, a subject's responses are treated as a k-dimensional response vector. Note that this approach is not as powerful as the univariate approach if the assumption of compound symmetry is accepted.

Unfortunately, neither the ε adjustments nor the multivariate approach protects the subtests that typically follow. This is particularly unfortunate since these subtests are used to clarify significant overall comparisons, and they are susceptible to bias under nonsphericity.

Difficulty in interpretation can occur when several dependent measures are made for each experimental treatment and the assumption of compound symmetry is rejected. This situation can result in a lack of degrees of freedom and power since the dimension of the response matrix, which is a multiple of the number of dependent variables and the number of unique within-subject factor treatment combinations, can equal or exceed the total number of subjects. In the multivariate context, this can result in the degrees of freedom parameter being very small. If the number of subjects is less than the number of unique within-subject treatment combinations, the overall multivariate test cannot be computed.

Since it is common and necessary to record, evaluate, and analyze numerous measurements during experimentation, alternate approaches to assess the effect of treatment conditions on the response measurements need to be explored. This report introduces and demonstrates the use of unbiased, efficient variance component estimators and their associated diagnostics to detect assumption violations concerning the repeated measures design. By identifying the source of the sphericity violation, the researcher would be able to modify the underlying linear model to account for the unexplained variability or remove outliers that may result in the acceptance of the sphericity assumption. Thus, a more powerful, simpler univariate repeated measures approach can be undertaken which may require a smaller sample size.

GENERAL VARIANCE COMPONENT ESTIMATES AND DIAGNOSTICS METHODOLOGY

The problem of estimating variance components in random and mixed models has been of interest to researchers for years as pointed out by Green and Hocking (1988). During the last few years, however, new closed form expressions for the estimators of variance components have been developed, based on the equivalence shown by Green (1985, 1987); Hocking (1985); and Hocking, Green, and Bremer (1989) of the variance component estimation problem to the problem of estimating the covariances, θ_t, between appropriately related observations. In addition, these estimators have been shown to provide information that will be useful in diagnosing problems and suggesting simple graphical procedures for examining the influence of the treatment levels.

To introduce this general methodology, this section only considers three-factor repeated measures design with Factors 1 and 3 repeated as shown in Table 1, with $\bar{Y}ijk.$ representing cell means. The number of levels of Factor (i) is designated by a_i. Subjects are designated Factor 2. Factors 1 and 3 are the within-subject fixed factors. The traditional univariate repeated measures model with subjects and subject interactions considered random is

$$Y(ijkm) = M+A(i)+S(j)+AS(ij)+B(k)+AB(ik)+SB(jk)+ABS(ijk)+\varepsilon(ijkm), \quad (1.1)$$

in which M is the overall mean, A(i) is the effect of level i of treatment or factor A, S(j) is the effect of subject j, AS(ij) is the effect of level ij of treatment combination AS, B(k) is the effect of level k of factor B, AB(ik) is the effect of the AB treatment combination at level ik, SB(jk) is the effect of treatment combination SB at level (jk), ABS(ijk) is the effect of level ijk of treatment combination ABS, and $\varepsilon(ijkm)$ is the random error. For the traditional univariate approach, it is assumed that A(i), B(k), AB(ik), and M are fixed, and S(j), AS(ij), SB(jk), ABS(ijk), $\varepsilon(ijkm)$ are zero mean, independent normal random variables with variances ϕ_2, ϕ_{12}, ϕ_{23}, ϕ_{123}, and ϕ_o, respectively. While the variables are independent, the responses are correlated with

$$\begin{aligned}
\theta_2 &= \phi_2, \text{ if } i \neq i^*, j = j^*, k \neq k^*, \\
\theta_{12} &= \phi_2 + \phi_{12}, \text{ if } i = i^*, j = j^*, k \neq k^*, \qquad (1.2) \\
\theta_{23} &= \phi_2 + \phi_{23}, \text{ if } i \neq i^*, j = j^*, k = k^*, \\
\theta_{123} &= \phi_2 + \phi_{12} + \phi_{23} + \phi_{123}, \text{ if } i = i^*, j = j^*, k = k^*, m^* \neq m,
\end{aligned}$$

$$\text{and } \theta_o = \phi_o + \theta_{123}, \text{ if } ijkm = i^*j^*k^*m^*.$$

This covariance structure suggests an alternate approach to the linear model first proposed by Hocking (1983) and extended and developed by Green (1985) to several classes of linear models. This approach relaxes the requirement that the variance components be positive. Thus, the classical model is replaced by specifying the response vector as normal with covariance matrix as given in equation (1.2) and mean vector determined from the expectation of Y, as

$$E(Y_{ijkm}) = M + A(i) + B(k) + AB(ik). \qquad (1.3)$$

The only restriction of the covariance matrix is that it must be positive definite. This requirement is weaker than the classical requirement that the ϕ_t be positive. An in-depth development of this alternate model is contained in Hocking (1985).

Table 1

Three-Factor Repeated Measures Design

FACTOR I

FACTOR III

SUBJECTS	1	2	. . .	a_1
	1 2 3 . . . a_3	1 2 3 . . . a_3	. . .	1 2 3 . . . a_3
1	$Y_{111}, Y_{112} \ldots Y_{11a_3}$	$Y_{211}, Y_{212} \ldots Y_{21a_3}$		$Y_{a_111}, Y_{a_112} \ldots Y_{a_11a_3}$
2	$Y_{121}, Y_{122} \ldots Y_{12a_3}$	$Y_{221}, Y_{222} \ldots Y_{22a_3}$		$Y_{a_121}, Y_{a_122} \ldots Y_{a_12a_3}$
3				
4				
.				
.				
.				
a_2	$Y_{1a_21}, Y_{1a_22} \ldots Y_{1a_2a_3}$	$Y_{2a_21}, Y_{2a_22} \ldots Y_{2a_2a_3}$		$Y_{a_1a_21}, Y_{a_1a_22} \ldots Y_{a_1a_2a_3}$

In this notation, θ_t is between observations at the same level of factors indexed by t and different levels of all other factors in the model. This suggests examining the corresponding sample covariances. These sample covariances yield the estimators of the θ_t. Sample covariances yielding estimators of θ_2 and θ_{12} are

$$\hat{\theta}_2 = \frac{1}{a_{13}\ r_{13}} \sum_{ik \neq i^*k^*} \frac{1}{r_2} \Sigma_j\ (\bar{Y}ijk. - \bar{Y}i.k.)(\bar{Y}i^*jk^*. - \bar{Y}i^*.k^*.), \text{ and} \qquad (1.4)$$

$$\hat{\theta}_{12} = \frac{1}{a_{13}\ r_3} \sum_{k \neq k^*} \frac{1}{r_2} \Sigma_{ij}\ (\bar{Y}ijk. - \bar{Y}i.k.)(\bar{Y}ijk^*. - \bar{Y}i.k^*.),$$

in which $\sum_{ik \neq i^*k^*}$ is the sum over $i \neq i^*$ and $k \neq k^*$ and $\sum_{ik}$ is the sum across all i and j. Similarly, θ_{23} is analogous to the θ_{12} estimator with subscripts i and k interchanged.

From equation (1.4), one recognizes the θ_2 estimator as the average of $a_{13} \times r_{13}/2$ equal expectation sample covariances corresponding to all combinations of $i \neq i^*$, $k \neq k^*$. As written, each sample covariance appears twice. Here, $r_i = a_i-1$. Similarly, θ_{12} is the average of $a_{13} \times r_3/2$ equal expectation sample covariances corresponding to all combinations of $i \neq i^*$ and $k \neq k^*$.

These covariances are unbiased and contain the diagnostic power. By plotting these covariances (diagnostics) in table form, one obtains an indication of the stability of the estimate and of suspect estimates.

For example, consider displaying the $a_{13} \times r_{13}$ sample covariances or diagnostics for θ_2, corresponding to $i \neq i^*$, $k \neq k^*$ into $a_1 \times r_1/2$ tables, each of dimension a_3 by a_3. In these tables, the off-diagonal terms are the distinct sample covariances associated with different levels of Factor 1 and Factor 3 and the same level of Factor 2. Also, for the three-factor design, consider θ_{12}. One displays the $a_1 \times a_3 \times r_3/2$ sample covariances in a_1 tables of dimension $a_3 \times a_3$. In these tables, the above diagonal elements are the $a_3 \times r_3/2$ distinct sample covariances associated with levels i of Factor 1 and level k of Factor 3, with all a_2 levels of Factor 2 used to determine the sample covariances.

In general, one looks for outliers and trends. For example, (a) unusually large or small diagonal entries indicate abnormal variability in the cell means for this level of the factor being investigated; (b) special patterns in the off-diagonal elements, such as a particular column or row having the majority of its entries higher or lower than other rows or columns, indicate that one or more cells may contain extreme outliers, or unexpected deviations from model assumptions occur; and (c) large fluctuations in the off-diagonal entries reflect high variability in the data.

Following the examination of the diagnostics, plots of treatment i versus treatment i* cell-means, in which abnormal diagnostics have been identified, are recommended. This will help the researcher identify the treatment cells responsible for extra large or small variance component estimates. Finally, the diagnostic procedure should conclude with an examination of the data in the identified cells.

REPEATED MEASURES DESIGN

To illustrate these diagnostic procedures, data from a repeated measures design performed by Malkin and Christ (1987) are used. The objective of the experiment was to conduct a laboratory flight simulation to compare a cockpit keyboard, a thumb-controlled switch, and a connected word voice recognizer for data entry of navigation map coordinate sets when (1) the entry of Universal Transverse Mercator (UTM) coordinate sets is the sole task performed (no flight), and (2) the entry of UTM coordinate sets is performed concurrently with controlling a helicopter simulator while flying a computer-generated external scene (flight). For this report, the differences among the three methods of data entry input time are evaluated for both the flight and no-flight conditions. Malkin and Christ also investigated subject error.

Methodology

Data were collected using 12 aviators assigned to Aberdeen Proving Ground, Maryland, as experimental subjects. The Aviation and Air Defense Division, Human Engineering Laboratory (HEL) flight simulator was used for this study. The crew simulator consists of a cockpit cab with advanced controls and displays and an "out-the-window" scene produced by computer-generated imaging (CGI). The CGI, cockpit controls, flight simulation, displays, and results were driven and recorded using two Digital Equipment Corporation (DEC) VAX® computers. All subjects were trained to operate the voice recognition system and flight simulator. For an in-depth account of the apparatus and training, refer to Malkin and Christ (1987).

Each subject entered eight UTM coordinate sets for each test condition. A standardized but unique set of coordinates was used in each condition. The subject was tested in both conditions using one data entry method before proceeding to the next data entry method. The order of the test conditions was counterbalanced to control for learning.

Experimental Design

A 2x3x12 factorial design with repeated measures on the 12 subjects was implemented. The within-subject factors were data entry methods (voice, keyboard, and thumb-controlled switch) and task conditions (flight, no flight). The dependent variables were input time and response time. For illustration purposes, the design, input time, and response time are contained in Tables 2 and 3, respectively.

Table 2

Method by Task by Subject
(mean input time)

	method					
	Voice		Keyboard		Thumb	
	flight	no flight	flight	no flight	flight	no flight
subject	1	2	1	2	1	2
1	15.8	17.8	16.9	16.8	28.5	34.3
2	23.9	49.3	9.1	13.2	25.0	35.5
3	33.0	55.9	13.6	31.6	29.7	48.8
4	15.2	27.8	11.3	16.1	24.1	43.1
5	35.9	45.0	11.9	20.7	39.2	65.2
6	49.8	36.4	11.8	23.7	36.3	49.1
7	27.2	34.9	13.9	20.6	31.7	44.7
8	20.6	20.6	10.9	24.1	35.4	37.4
9	28.9	38.7	10.5	19.9	34.7	34.6
10	27.7	23.5	10.7	15.9	34.0	43.6
11	17.9	11.7	15.4	24.1	32.6	39.0
12	23.0	16.3	13.5	33.8	38.9	70.9

Table 3

Method by Task by Subject
(mean response time)

	method					
	Voice		Keyboard		Thumb	
	Task 1		Task 1		Task 3	
	no flight	flight	no flight	flight	no flight	flight
subject	1	2	1	2	1	2
1	2.03	1.91	1.02	3.60	0.64	0.54
2	2.03	3.10	1.02	2.92	5.44	9.10
3	2.53	5.90	0.76	4.69	5.94	11.61
4	1.77	3.85	1.15	4.86	1.26	2.28
5	1.88	5.11	0.77	9.93	1.16	3.64
6	2.46	4.00	1.77	14.31	3.03	6.67
7	1.89	2.10	1.52	9.29	1.14	2.14
8	1.64	1.36	1.01	1.37	1.02	0.47
9	2.53	2.26	1.02	5.59	1.02	1.77
10	2.02	3.45	0.89	4.13	0.72	3.32
11	1.65	1.40	1.22	5.96	0.76	1.94
12	3.67	6.90	2.15	7.30	1.01	8.72

Results

Since the response measures were highly correlated and only 12 subjects were used, a multivariate analysis of variance (MANOVA) was performed with task and method fixed and subjects considered a random factor. This approach is suggested by Schutz and Gessorali (1987). The approximate F-ratios were then checked against the Greenhouse and Geisser (1959) adjustment, and they agreed.

The results are shown in Figure 1, with approximate F presented based on Wilk's lambda. For response time, subjects were able to respond significantly faster during the no-flight condition than during the flight condition. There also was a significant interaction between data entry method and task conditions. During the no-flight task condition, subjects responded significantly faster when the keyboard was used to enter data. During the flight task condition, however, subjects responded significantly faster using either voice or the thumb-controlled switch (see Figure 2).

There were significant differences among the three mean input times for the data entry method. Subjects were able to input data faster during the no-flight task conditions than during the flight conditions. However, there was no significant interaction between task and entry method (see Figure 3).

As a final note, the input time covariances for the within-subject factors showed extreme deviation from the sphericity assumption, whereas the sphericity for response time was acceptable based on Mauchly's (1940) test criteria for the task and method main effects, but the sphericity assumption was rejected for the task by method interaction. The variance component diagnostic procedure is demonstrated for input times and response time. An in-depth analysis for input time is provided, which demonstrates the use of these diagnostics for checking and identifying assumption violations. The diagnostics for response time are also presented.

ILLUSTRATED EXAMPLE OF VARIANCE COMPONENT ESTIMATES AND DIAGNOSTICS

As previously noted, it is natural to estimate the covariances θ_t by corresponding sample covariances. In the balanced case and for the Malkin and Christ data, the estimates can be obtained from the analysis of variance (ANOVA) table by equating mean squares to expected mean squares (see Figure 4). Based on this method, the estimates for input times were calculated as $\hat{\phi}_2 = 7.68$, $\hat{\phi}_{12} = 31.03$, $\hat{\phi}_{23} = 3.35$, $\hat{\phi}_{123} = -5.49$, $\hat{\phi}_o = 305.26$, and for response time, $\hat{\phi}_2 = 0.523$, $\hat{\phi}_{12} = 1.86$, $\hat{\phi}_{32} = 1.38$, $\hat{\phi}_{123} = 1.16$ $\hat{\phi}_o = 8.04$.

For this example, $a_1 = 3$, $a_2 = 12$, and $a_3 = 2$. The estimate of θ_2 is the average of six distinct sample covariances. They can be displayed in a table such as Table 4 for input time or Table 5 for response time. The elements are the sample covariances. To avoid confusion, it is worth noting that the diagonal elements are not true variances since $k \neq k^*$.

Under the covariance structure and compound symmetry assumption, all off-diagonal elements of Tables 4 and 5 should be approximately equal as well

DEPENDENT MEASURES	F-STATISTICS	SUBJECT SIGNIFICANCE = .05
RESPONSE TIME	6.64	•
INPUT TIME	5.57	•

DEPENDENT MEASURES	F-STATISTICS	METHOD SIGNIFICANCE = .05
RESPONSE TIME	1.07	
INPUT TIME	30.78	•

DEPENDENT MEASURES	F-STATISTICS	TASK SIGNIFICANCE = .05
RESPONSE TIME	29.25	•
INPUT TIME	25.21	•

DEPENDENT MEASURES	F-STATISTICS	METHOD BY TASK SIGNIFICANCE = .05
RESPONSE TIME	11.7	•
INPUT TIME	2.25	

DEPENDENT MEASURES	F-STATISTICS	SUBJECT BY METHOD BY TASK SIGNIFICANCE = .05
RESPONSE TIME	1.13	
INPUT TIME	1.00	

DEPENDENT MEASURES	F-STATISTICS	SUBJECT BY TASK SIGNIFICANCE = .05
RESPONSE TIME	4.09	•
INPUT TIME	2.45	•

DEPENDENT MEASURES	F-STATISTICS	SUBJECT BY METHOD SIGNIFICANCE = .05
RESPONSE TIME	3.15	•
INPUT TIME	2.09	•

Figure 1. MANOVA results for three-factor repeated measures design with response time and input time the dependent measure.

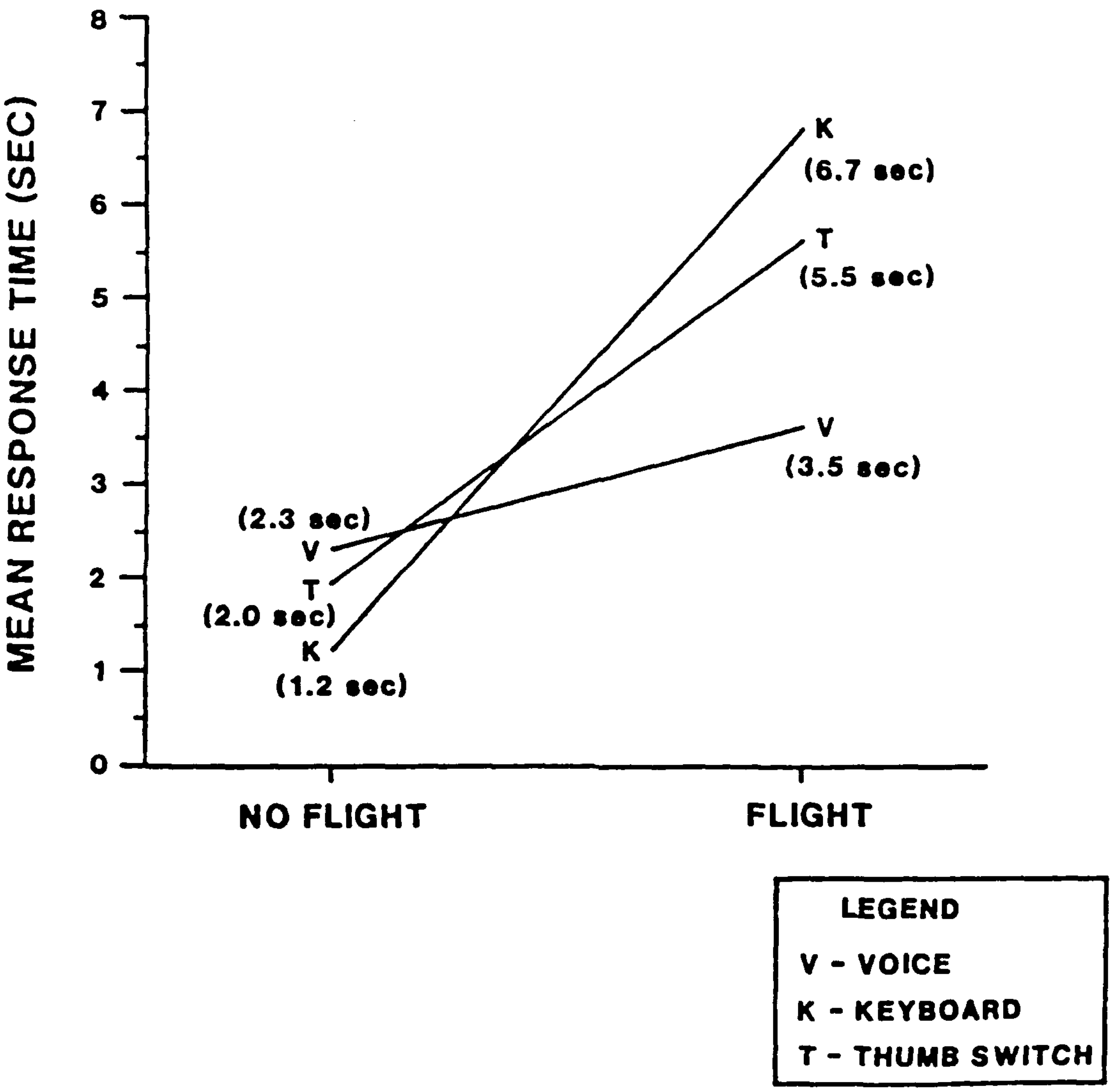

Figure 2. Data entry methods by task for response time.

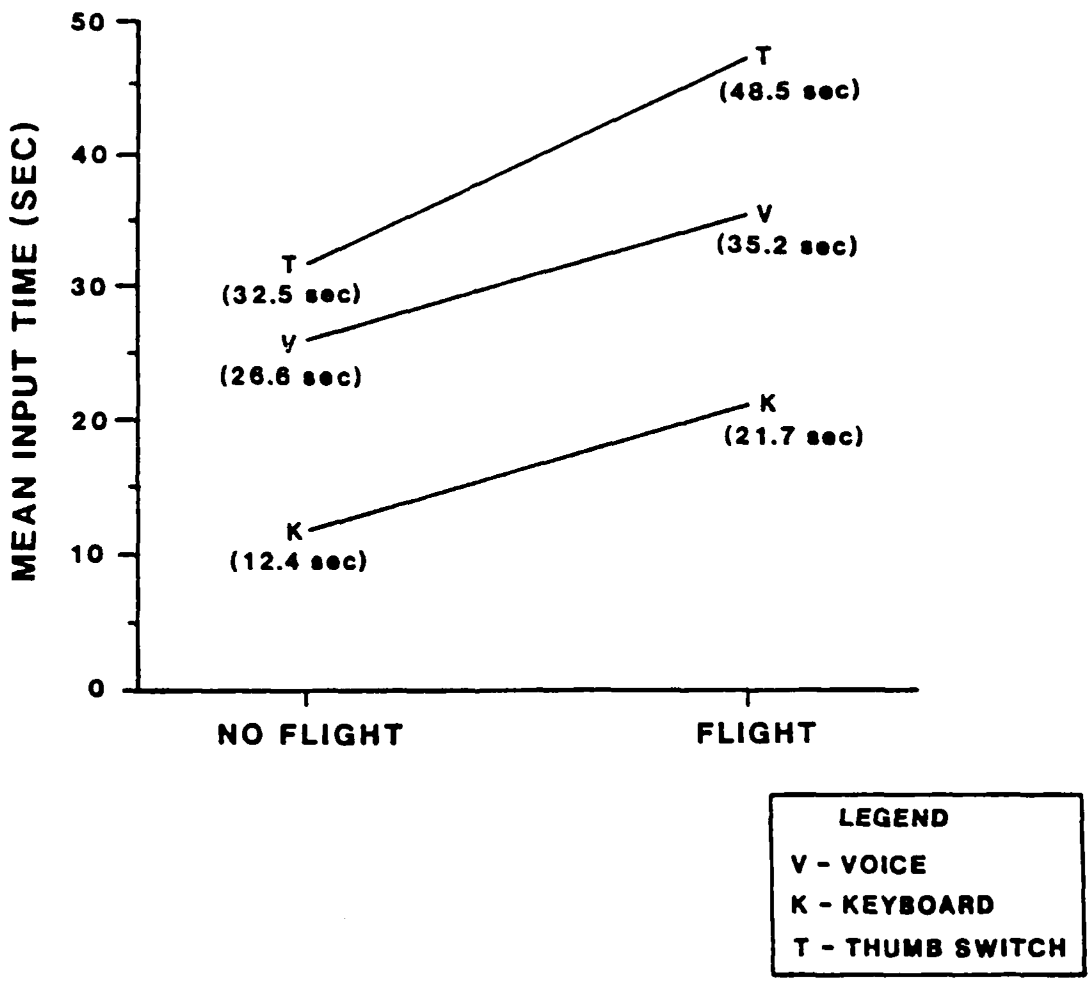

Figure 3. Data entry methods by task for input time.

SOURCE	dF	E M S
METHOD	2	$\theta_o+n\phi_{123}+na_3\phi_{12}+na_2a_3M_1$
TASK	1	$\theta_o+n\phi_{123}+na_1\phi_{23}+na_1a_2T_3$
METH x TASK	2	$\theta_o+n\phi_{123}+na_2MT_{13}$
SUB	11	$\theta_o+na_1a_3\phi_2+na_3\phi_{12}+n\phi_{123}$
SUB x METHOD	22	$\theta_o+na_3\phi_{12}+n\phi_{123}$
SUB x TASK	11	$\theta_o+na_1\phi_{23}+n\phi_{123}$
SUB x METHOD x TASK	22	$\theta_o+n\phi_{123}$
ERROR	504	θ_o

Figure 4. Analysis of variance for the three-way repeated measures model. (Method and task are within-subject factors; subjects are random.)

Table 4

Diagnostics Input Time
(all subjects included)

θ_2

NO FLIGHT

		VOICE	KEYBOARD	THUMB
FLIGHT	VOICE	76.20	-13.80	-12.47
	KEYBOARD	13.68	4.80	16.01
	THUMB	40.78	1.88	35.10

Table 5

Diagnostics Response Time
(all subjects included)

θ_2

NO FLIGHT

		VOICE	KEYBOARD	THUMB
FLIGHT	VOICE	0.72	0.15	1.22
	KEYBOARD	0.51	0.77	-0.26
	THUMB	1.26	0.24	5.72

as all diagonal elements. Therefore, the diagnostics provide an illustrative procedure to check the compound symmetry assumption and identify treatment combinations that indicate violations of this assumption.

In examining the θ_2 off-diagonal diagnostics of Table 4, the covariances keyboard no flight versus voice flight (-13.80) and thumb no flight versus voice flight (-12.47) are small when compared to the other off-diagonal entries in the table. In addition, thumb flight versus voice no flight (40.78) seems large in comparison. This large fluctuation indicates high variability in the data.

Further examination of the specified treatment combinations is suggested. Follow-up plots of subject mean input times by treatment combinations, reflecting the large or small covariances, are shown in Figures 5 and 6. Examination of these plots reveals that for subjects 3, 5, 6, and 12, input time contributed to the extremely high or low covariances.

No specific patterns or large fluctuations in the off-diagonal elements were prevalent in the θ_2 response time diagnostics of Table 5. The diagonal element thumb no flight versus thumb flight seemed large compared to the other diagnostics. It is interesting to note that Mauchly's criteria for method had an approximate chi-square of 5.8 with 2 degrees of freedom, which was not significant at the .05 level. However, Mauchly's test for task by method was significant at the .05 level, and the assumption of compound symmetry was rejected. A plot of the subject response cell means for thumb no flight versus thumb flight (see Figure 7) revealed that subjects 6, 3, and 2 contributed to this large diagonal element, and subject 12 was also atypical from the rest of the subjects.

The diagnostic plots for θ_{12} and θ_{23} are *shown* in Tables 6 and 7. For θ_{12}, the plot consists of covariances based on the same level of subject and method but different levels of task. The diagnostic plot for input time revealed a large covariance component of 76.2 for voice no flight versus voice flight and a small covariance of 4.8 for keyboard no flight versus voice flight. A follow-up plot for the large covariance (see Figure 7) indicated that subjects' 3, 5, and 6 input times contributed to this large covariance.

The response time diagnostic plot (see Table 7) for θ_{12} reveals that the off-diagonal element thumb flight versus thumb no flight was large (5.72) compared to the other off-diagonal element as previously stated. Subjects 2, 3, 6, and 12 contributed to this diagnostic's magnitude as shown in Figure 8.

Similarly, the diagnostic plot for θ_{23} reveals large spurious covariances at treatment combinations voice no flight versus thumb no flight (23.1) and keyboard flight versus thumb flight (43.2) for input time. Note that this diagnostic plot contains covariances based on the same subject and task levels but different methods. Follow-up plots based on input time (see Figures 9 and 10) for both covariances reveal that subjects' 3, 5, 6, and 12 input times were contributing to one or both large covariance components.

Diagnostic plots for θ_{23} revealed no spurious covariances as shown in the off-diagonal elements of Table 6. However, for the no-flight condition, the diagonal element thumb was substantially larger when compared to the other

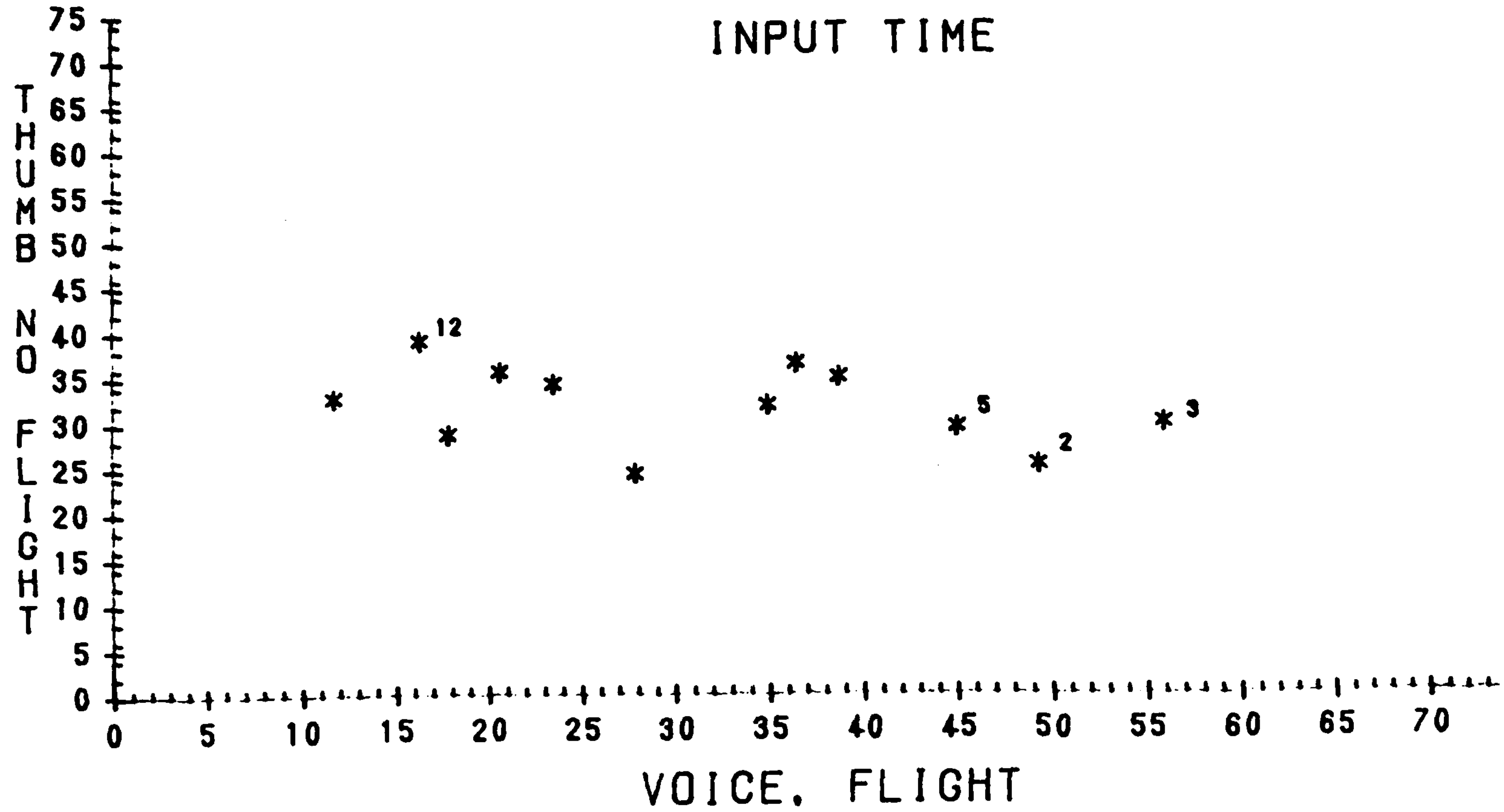

Figure 5. Subject cell means of thumb no flight versus voice flight. (Plotting symbol: Numbers represent a distinct subject.)

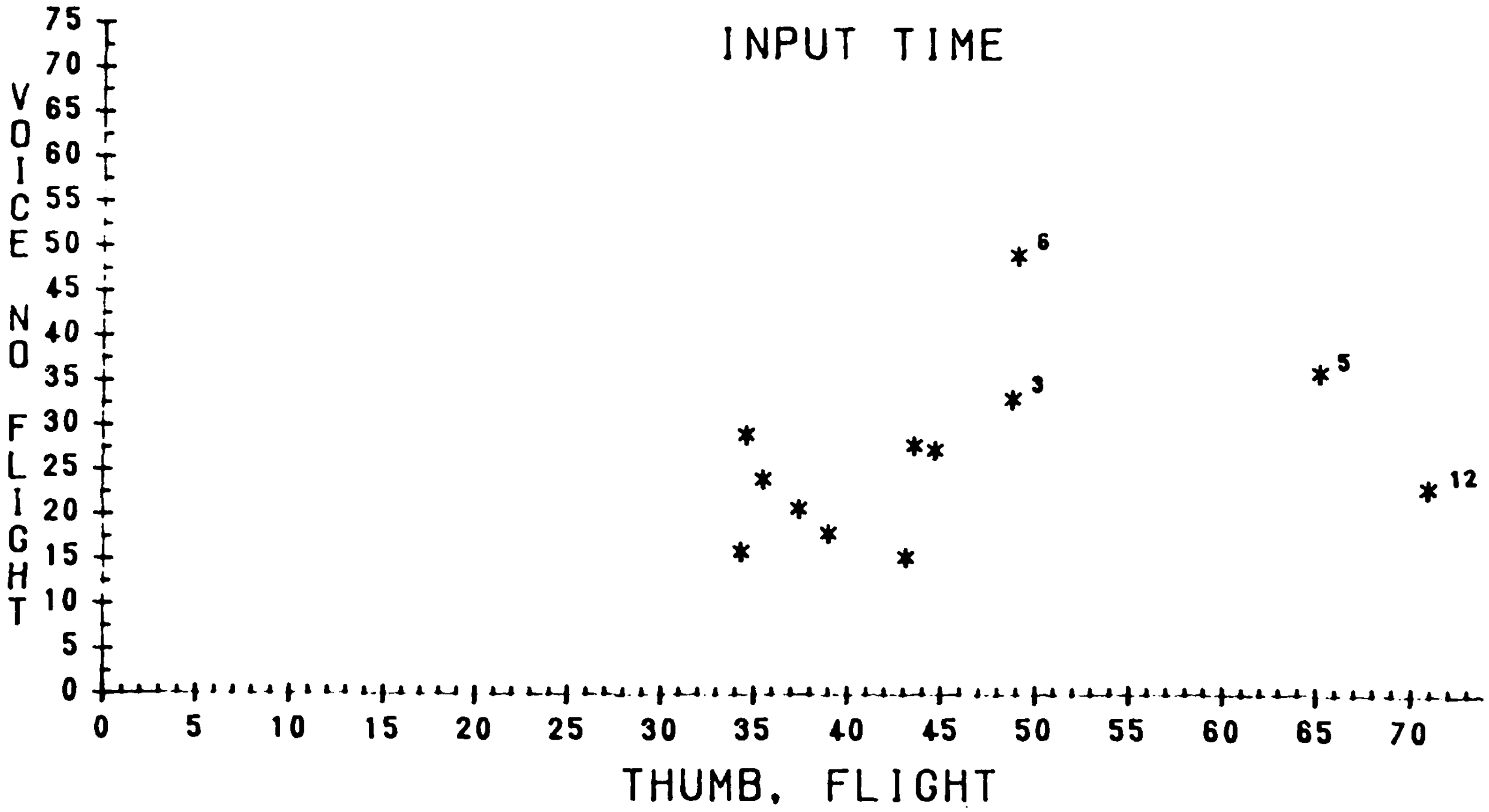

Figure 6. Subject cell means of voice no flight versus thumb flight. (Plotting symbol: Numbers represent a distinct subject.)

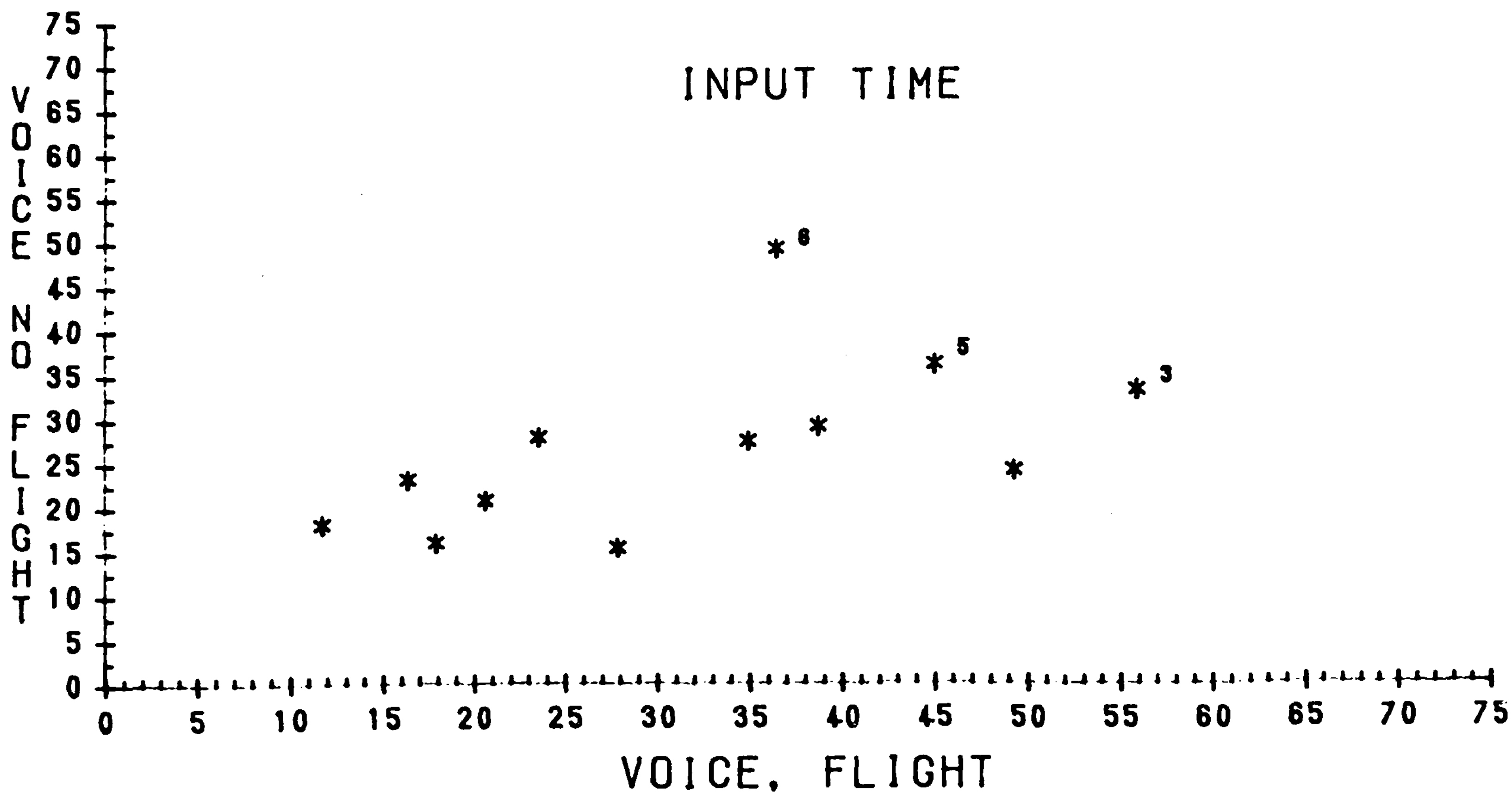

Figure 7. Subject cell means of voice no flight versus flight. (Plotting symbol: Numbers represent a distinct subject.)

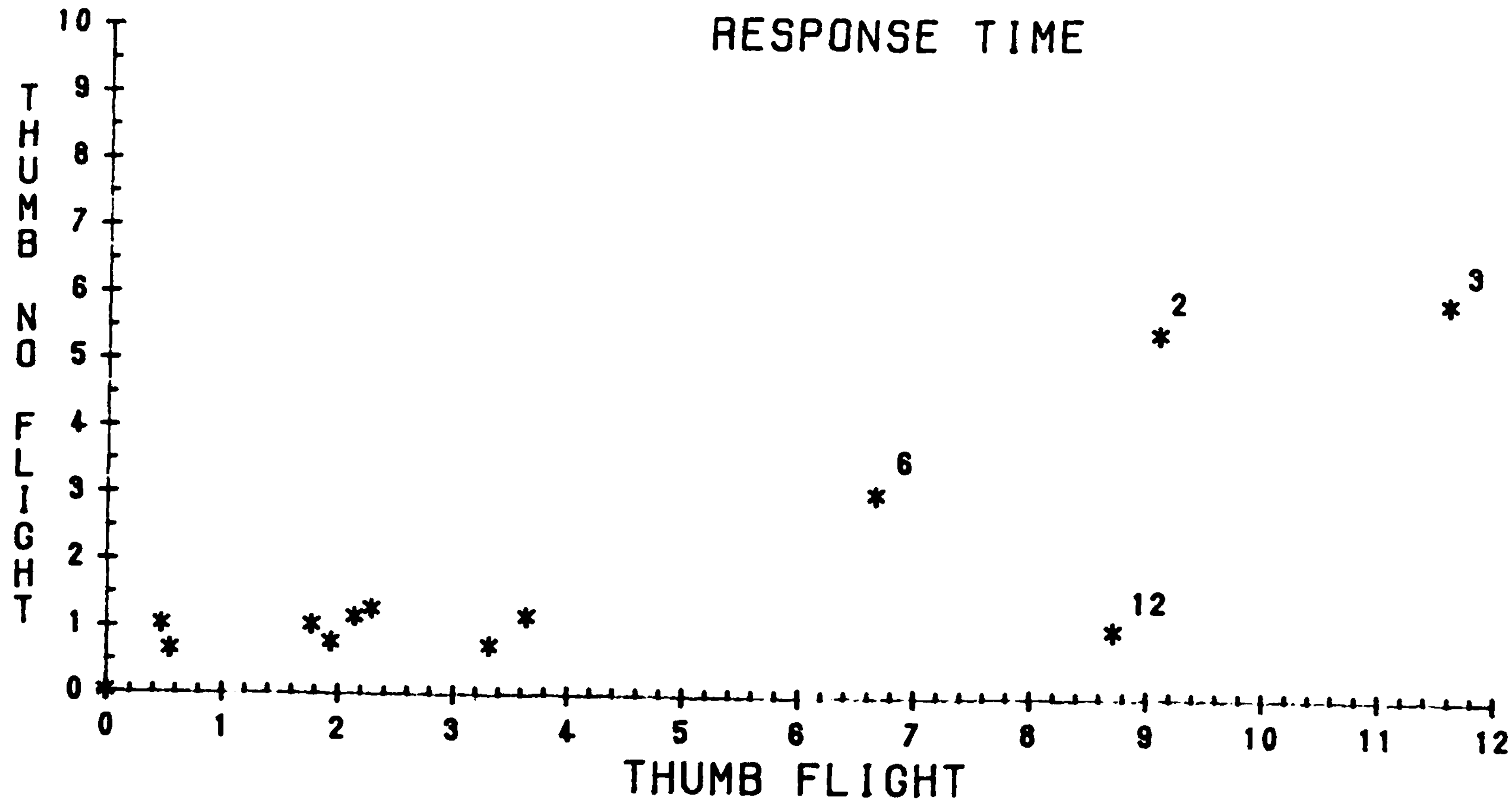

Figure 8. Subject cell means of voice no flight versus voice flight. (Plotting symbol: Numbers represent a distinct subject.)

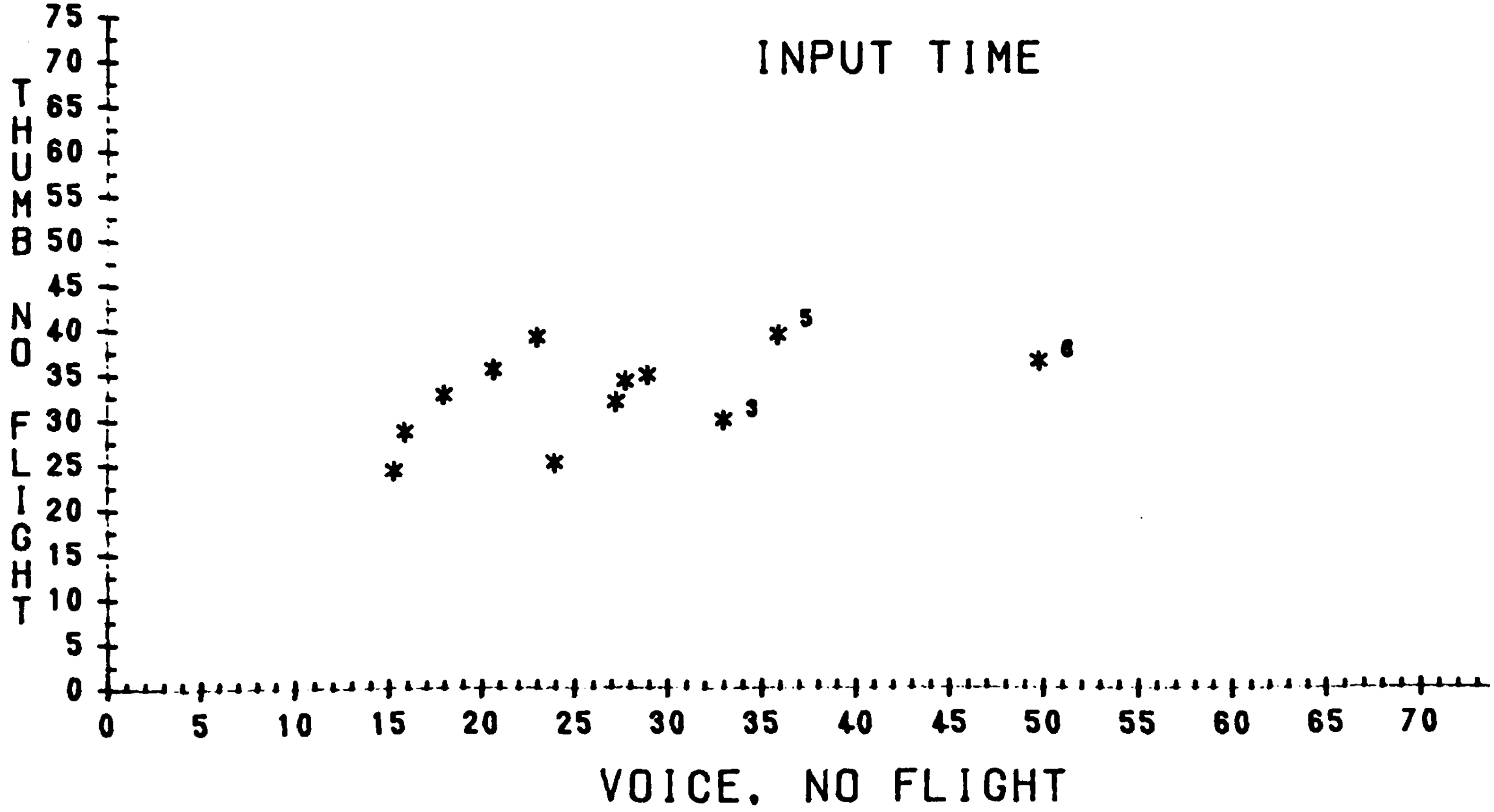

Figure 9. Subject cell means of voice no flight versus thumb no flight. (Plotting symbol: Numbers represent a distinct subject.)

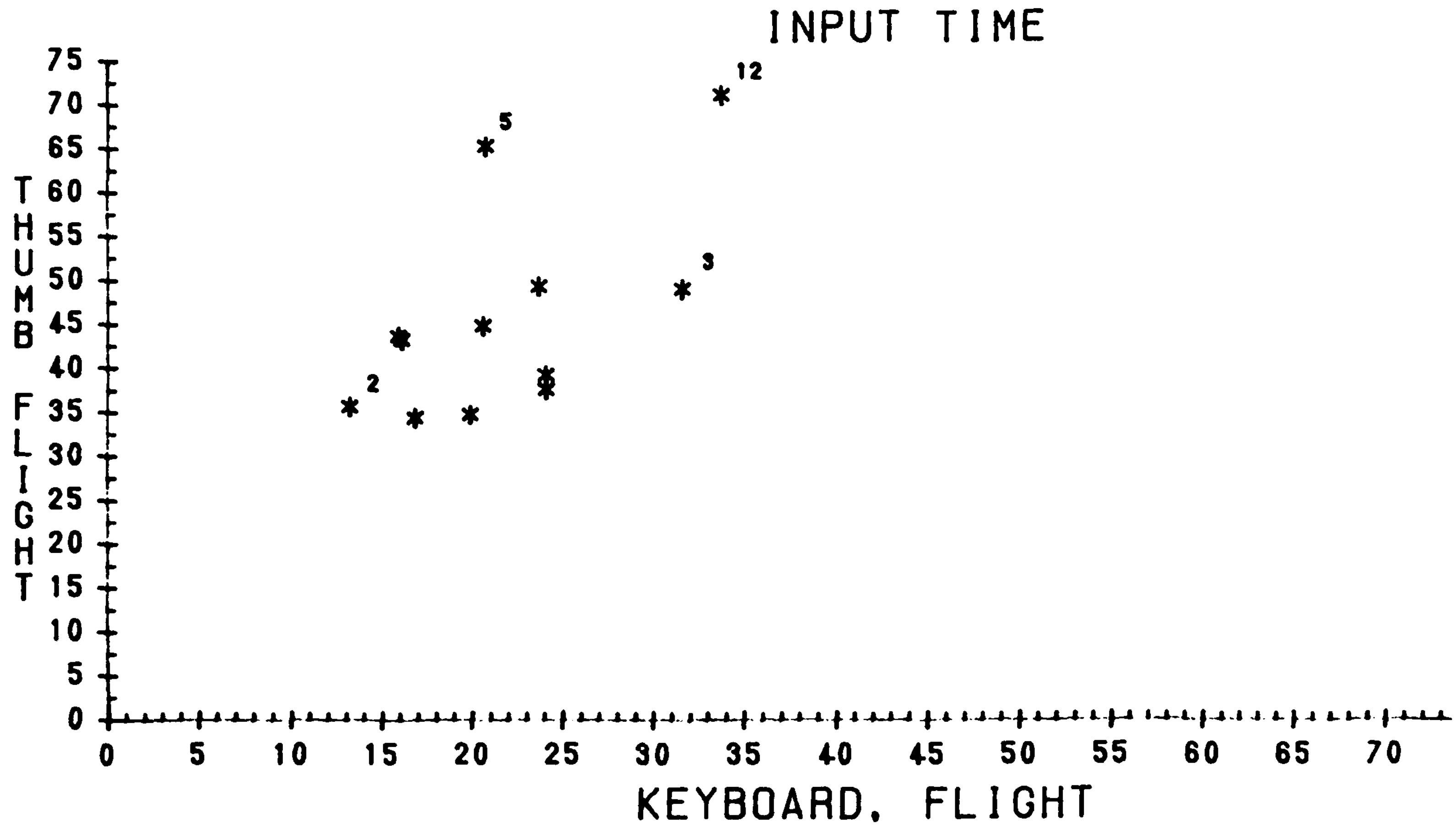

Figure 10. Subject cell means of voice no flight versus thumb no flight. (Plotting symbol: Numbers represent a distinct subject.)

Table 6

Diagnostics Input Time
(all subjects included)

Θ_{12}

METHOD (l)

1*=1	VOICE TASK (k)		KEYBOARD TASK (k)		THUMB TASK (k)	
	NO FLIGHT (1)	FLIGHT (2)	NO FLIGHT (1)	FLIGHT (2)	NO FLIGHT (1)	FLIGHT (2)
NO FLIGHT	95.3	76.2	5.1	4.8	24.09	35.1
FLIGHT	76.2	19.9	4.8	38.7	35.1	138.1

Θ_{23}

TASK (k)

	NO FLIGHT METHOD (l)			FLIGHT METHOD (l)		
	VOICE	KEYBOARD	THUMB	VOICE	KEYBOARD	THUMB
VOICE	95.3	-5.9	23.1	199.0	-5.0	10.5
KEYBOARD	-5.9	5.1	0.19	-5.0	38.7	43.2
THUMB	23.1	0.19	24.1	10.5	43.3	138.2

Table 7

Diagnostics Response Time
(all subjects included)

Θ_{12}

METHOD

	VOICE TASK		KEYBOARD TASK		THUMB TASK	
	NO FLIGHT (1)	FLIGHT (2)	NO FLIGHT (1)	FLIGHT (2)	NO FLIGHT (1)	FLIGHT (2)
NO FLIGHT	0.32	0.72	0.19	0.76	3.47	5.72
FLIGHT	0.72	3.22	0.76	12.73	5.72	13.87

Θ_{23}

TASK

NO FLIGHT METHOD

	VOICE	KEYBOARD	THUMB
VOICE	0.32	0.13	0.16
KEYBOARD	0.13	0.19	-0.18
THUMB	0.16	-0.18	3.45

FLIGHT METHOD

	VOICE	KEYBOARD	THUMB
VOICE	3.22	2.08	5.09
KEYBOARD	2.08	12.73	2.16
THUMB	5.09	2.16	13.87

diagonal elements. For the flight condition, the voice diagonal was smaller. Identifying what seemed to be a dichotomous population of subjects, a review of subject records was undertaken to try to explain the reason that several subjects seemed to respond differently than the others. A review of the records indicated that in general, these pilots were older (over 42 as compared to under 38), had a higher military rank, and had spent as much time or more flying fixed wing as rotary wing aircraft, with recent flying experience concentrated on fixed wing. Subject 2's records on demographics were missing. Based on this response, however, the pilot will be considered part of the fixed wing group. Based on subjective input from experienced pilots, differences between the aircraft in regard to instrumentation and flying procedures could certainly account for the difference in input times between fixed wing and rotary wing pilots.

A recalculation of the diagnostics for input time with subjects 2, 3, 5, 6, and 12 removed revealed covariances that were more stable as shown in Tables 8 and 9. Regarding response time, a recalculation of the diagnostics with subjects 2, 3, 5, 6, and 12 also demonstrated more stable diagnostics as shown in Tables 10 and 11.

It is wor.h noting that Mauchly's approximate chi-square for input time was not significant at the .05 level for method or for the method by task interaction. The value of the test statistic was 2.16 with 2 degrees of freedom and 1.78 with 2 degrees of freedom, respectively. This trend was also observed for response time. The chi-square statistics were 4.47 and 5.06 with 2 degrees of freedom.

Alternatively, if the subjects are grouped into fixed wing and rotary wing categories and the data are analyzed as a four-factor experiment, the assumption of compound symmetry should be more plausible. The between-subject group category will be included in the linear model and the data re-analyzed based on this modified model at a later time in conjunction with the newly developed distribution theory.

CONCLUSIONS AND RECOMMENDATIONS

The variance component estimates and associated diagnostic procedures have been shown to be computationally and intuitively simple. All calculations can be obtained using standard statistical packages such as SPSSX$^{X™}$, SAS®, or BMDP. For example, the procedure CANDISC in SAS® will calculate the desired covariances.

The diagnostic procedures have been demonstrated to be effective in checking the assumptions underlying the repeated measures model and useful in identifying probable causes for the violation of these assumptions. Thus, an alternate approach is proposed that provides the researcher the option of removing spurious observations, performing transformations, or controlling additional sources of variability so that the data can conform to the standard assumptions or to modifying the model. By circumventing the problems associated with the traditional univariate repeated measures analysis, these diagnostic procedures provide an easier interpretation of the results and increased validity of the conclusions derived from the data. The result is a

Table 8

Diagnostics Input Time
(subjects 2, 3, 5, 6 and 12 deleted)

Θ_2

NO FLIGHT

		VOICE	KEYBOARD	THUMB
FLIGHT	VOICE	37.15	0.48	24.07
	KEYBOARD	-13.01	0.86	-3.78
	THUMB	1.93	8.11	4.84

Table 9

Diagnostics Input Time
(subjects 3, 5, 6 and 12 deleted)

θ_{12}

METHOD

	VOICE TASK		KEYBOARD TASK		THUMB TASK	
	NO FLIGHT (1)	FLIGHT (2	NO FLIGHT (1)	FLIGHT (2)	NO FLIGHT (1)	FLIGHT (2)
NO FLIGHT	35.03	37.15	6.80	0.86	15.96	4.85
FLIGHT	37.15	90.68	0.86	12.61	4.85	39.66

θ_{23}

TASK

NO FLIGHT METHOD

	VOICE	KEYBOARD	THUMB
VOICE	35.02	-7.48	16.06
KEYBOARD	-7.48	6.80	-2.95
THUMB	16.06	-2.95	15.96

FLIGHT METHOD

	VOICE	KEYBOARD	THUMB
VOICE	90.68	-8.50	37.63
KEYBOARD	-8.50	12.61	3.26
THUMB	37.63	3.26	39.66

Table 10

Diagnostics Response Time
(subjects 2, 3, 5, 6 and 12 deleted)

Θ_2

NO FLIGHT

		VOICE	KEYBOARD	THUMB
FLIGHT	VOICE	0.11	0.04	-0.10
	KEYBOARD	-0.14	0.42	0.16
	THUMB	0.08	0.22	-0.02

Table 11

Diagnostics Response Time
(subjects 2, 3, 5, 6 and 12 deleted)

θ_{124}

METHOD

	VOICE TASK		KEYBOARD TASK		THUMB TASK	
	NO FLIGHT (1)	FLIGHT (2	NO FLIGHT (1)	FLIGHT (2)	NO FLIGHT (1)	FLIGHT (2)
NO FLIGHT	0.11	0.07	0.21	0.42	0.05	0.02
FLIGHT	0.07	0.11	0.42	5.87	0.02	3.51

θ_{234}

TASK

NO FLIGHT METHOD

	VOICE	KEYBOARD	THUMB
VOICE	0.10	-0.03	0.01
KEYBOARD	-0.03	0.07	0.02
THUMB	0.01	0.01	0.04

FLIGHT METHOD

	VOICE	KEYBOARD	THUMB
VOICE	0,79	0.27	0.27
KEYBOARD	0.27	5.87	1.17
THUMB	0.27	1.17	3.51

DIAGNOSTIC D

valuable statistical approach that can be applied in many areas including medicine, physiology, agriculture, and quality control.

FUTURE RESEARCH

The covariance structure of the response vector, as indicated in equation (1.2), suggests an alternate approach to modeling and analyzing random effects. First introduced by Hocking (1983), it was extended to designs of all sizes and types by Green (1985, 1987) and Hocking, Green, and Bremer (1989). The problem of estimating the individual variance components is equivalent to the problem of estimating covariances using sample covariances. For example, to estimate ϕ_1 in a three-way random crossed design, one can use

$$\frac{1}{(a_1-1)} \times \sum_{i=1}^{a_1} (\bar{Y}ijk. - \bar{Y}.jk.)(\bar{Y}ij^*k^*. - \bar{Y}.j^*k^*.)$$

for each combination $j \neq j^*$, $k \neq k^*$, $\bar{Y}ijk. = \frac{1}{n_{ijk}} \times \sum_{m=1}^{n_{ijk}} Yijkm$, and

$$\bar{Y}.jk = \frac{1}{a_1} \times \sum_{i=1}^{a_1} \bar{Y}ijk..$$

The dot representing the cell mean, based on repetition, will be dropped for simplicity when it is clear that $\bar{Y}ijk.$ is a cell mean.

Equivalently, one can write the covariance as

$Z'_1AZ_2/(a_1-1)$ in which

$Z'_1 = (\bar{Y}1jk., \bar{Y}2jk., \ldots, \bar{Y}a_1jk.) = (\bar{Y}ijk.)_i$,

$Z'_2 = (\bar{Y}1j^*k^*., \bar{Y}2j^*k^*., \ldots, \bar{Y}a_1j^*k^*.) = (\bar{Y}ij^*k^*.)_i$, and

$A = I_{a_1} - J_{a_1}J_{a_1}'/a_1$,

with I an identity matrix and J_{a_1} a vector of ones of length a_1. Z'_1AZ_2 is referred to in the literature as a bilinear form.

The covariance is calculated for the same level of factor one and different levels of factors two and three. Here, a_i is the number of levels of the i-th factor. These $(a_2)(a_2-1)(a_3)(a_3-1)/2$ covariances or diagnostic components are averaged to obtain an unbiased estimate of ϕ_1, called the "ave" estimator of ϕ_1.

These covariances provide the diagnostic power for examining the quality of the variance component estimates, isolating outliers, and identifying model deficiencies. This diagnostic methodology needs to be formalized by developing the distribution of the bilinear form. In doing so, the ability to determine the probability of obtaining a covariance of a specific magnitude for a particular linear model will be accomplished, and a metric to formally identify problems with a random or mixed linear model will be established.

To accomplish this, two separate cases that depend on the variance-covariance structure of $Z' = (Z'_{1.}, Z'_{2.})$ need to be considered. In the first case, the covariance structure of Z is

$$V = \begin{pmatrix} aI & bI \\ bI & aI \end{pmatrix}$$

Thus, the paired observations comprising the bilinear form are independent. For $i \neq i^*$, (Z_{1i}, Z_{2i}) and (Z_{1i^*}, Z_{2i^*}) are independent. In the second case, the variance-covariance structure of Z is

$$V = \begin{pmatrix} aI + bJJ' & cI + dJJ' \\ cI + dJJ' & aI + bJJ' \end{pmatrix}$$

The paired observations comprising the bilinear form are dependent. For $i \neq i^*$, (Z_{1i}, Z_{2i}) and (Z_{1i^*}, Z_{2i^*}) are dependent.

Thus, the distribution of this set of samples will be developed in both cases, thereby giving a formal basis for a diagnostic procedure that has been demonstrated useful in identifying deficiencies. This will result in the generation of comprehensive critical tables for the percentiles .01, .05, .10, .90, .95, and .99 of the distribution for a range of correlations and sample sizes. In addition, a computer program, which will generate the critical values regardless of the sample size or correlation, needs to be written. Thus, a mixed or random analog of residual analysis, complete with diagnostic tools, will be developed.

REFERENCES

Green, J. W. (1985). _Variance components: Estimates and diagnostics._ Doctoral dissertation, Texas A & M University.

Green, J. W. (1987). _Diagnostic procedure for variance components for both large and small designs._ Manuscript submitted for publication.

Green, J. W., & Hocking, R. R. (1988). Model based diagnostics for variance components in a general mixed linear model, _Proceedings of the Thirty-Third Conference on the Design of Experiments._ Research Triangle Park, NC: U.S. Army Research Office.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. _Psychometrika_, _24_, 95-112.

Hocking, R. R. (1983). A diagnostic tool for mixed models with application to negative estimates for variance components. _Proceedings of the Eighth Annual SAS Users Group International Conference_ (pp. 711-716). New Orleans, LA.

Hocking, R. R. (1985). _The analysis of linear models._ Monterey, CA: Publishing Company.

Hocking, R. R., Green, J. W., & Bremer, R. H. (1989). Estimation of fixed effects and variance components in mixed factorial models including model-based diagnostics. _Technometrics_, _31_, 227-239.

Huynh, H., & Feldt, L. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. _Journal of the American Statistical Association_, _65_, 1582-1589.

Malkin, F. J., & Christ, K. A. (1987). _Comparisons of alphanumeric data entry methods for advanced helicopter cockpits_ (Technical Memorandum 14-87). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory.

Mauchly, J. W. (1940). Significance tests for sphericity of a normal _n_-variate distribution. _Annals of Mathematical Statistics_, _11_, 204-209.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. _Psychological Bulletin_, _97_, 316-333.

Schultz, R. W., & Gessaroli, M. E. (1987). The analysis of repeated measures designs involving dependent variables. _Research Quarterly for Exercise and Sport_, _58_, 132-149.