ANIMAL scientists often analyze data recorded on several occasions for each animal under different experimental conditions (i.e., non-random sampling). In a well-planned experiment these conditions are relatively homogeneous except for one property, which the experimenter wishes to examine. Commonly, the time-scale is of interest, as in pregnancy, lactation or growth. In other cases, information is obtained after each of several different treatments have been applied sequentially to the same animal, but the various sequences are not balanced as in changeover designs (Cochran and Cox, 1957).

The structure of the underlying causes of variation often is over-simplified in analysis of such data. Simplifications usually fall into one or two categories; 1) Analyzing factorial experiments as if completely randomized, when the structure is really split-plot (i.e., Sokal and Rohlf, 1969) and 2) Ignoring the correlations of errors induced by repeated measurement—although this constitutes a serious mistake only if the correlation is not uniform over the time-scale observed. The first type of mistake leads to distorted probabilities in making inferences about either repeated or non-repeated factors; the second often affects only the repeated factor (i.e., time).

The objective of this paper is to present statistical procedures that should be used to avoid grossly misleading inferences. This will be done by example, with certain generalities and computations committed to an appendix.

Example

The data in table 1 (Paape and Tucker, 1969) were selected for ease of presentation, and are only representative of a class of experiments. The simplified problem concerns the influence of pregnancy upon concurrent lactational performance of rats as measured by litter weight gains, where litter size was adjusted to maintain one suckling pup per mammary gland, and litters were replaced every 4 days to maintain intense suckling stimulus.

Analysis as Completely Randomized. If one makes the simplifying assumption that this two-factor experiment can be analyzed as if it were completely randomized (Snedecor and Cochran, 1967) i.e., as if the six observations within any treatment-period combination were independent of those in any other cell, the analysis of variance emerges as shown in table 2. Inferences from this analysis indicate the effects of pregnancy and time are highly significant, and the two factors apparently do not interact to a large degree. However, one may feel intuitively that lack of mutual independence among the cells of data could have distorted the results (because of repeated measurements on the rats).

Analysis as a Split-Plot. The split-plot nature of the experiment becomes evident (Snedecor and Cochran, 1967) if one recognizes that comparisons across treatment groups are free to vary more (variation between rats) than are comparisons across periods (variation within rats). Then the analysis proceeds as shown in table 3. The inferences are largely inconsistent with the previous analysis. In the split-plot analysis, the effect of time is highly significant, but the effect of pregnancy is not. In further contrast to the previous analysis, the interaction is highly significant, indicating that the time-trend in litter weight gains differs for pregnant and non-pregnant rats. Inferences are altered because variation between rats is increased by positive correlation of repeated observations; whereas, variation within rats is decreased relative to random variation.

In addition to the usual assumption of equal variances, the validity of comparisons within rats (i.e., among periods) depends on the uniformity of correlations between data for any two periods. Statistically, this means a homogeneous variance-covariance matrix. Experience indicates one would be well-advised to examine this assumption for data from experiments with non-random repeated measurements, for it is often true that the correlation of data from two periods declines with larger difference in time. However, Geisser and Greenhouse (1958) suggested a procedure
TABLE 1. LITTER WEIGHT GAINS (g) FOR PREGNANT AND NON-PREGNANT LACTATING RATS

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Rat number</th>
<th>Period of lactation (days)</th>
<th>Avg</th>
<th>8-12</th>
<th>12-16</th>
<th>16-20</th>
<th>20-24</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pregnant</td>
<td>1</td>
<td>7.5</td>
<td>8.6</td>
<td>6.9</td>
<td>0.8</td>
<td>5.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10.6</td>
<td>11.7</td>
<td>8.8</td>
<td>1.6</td>
<td>8.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>12.4</td>
<td>13.0</td>
<td>11.0</td>
<td>5.6</td>
<td>10.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>11.5</td>
<td>12.6</td>
<td>11.1</td>
<td>7.5</td>
<td>10.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>8.3</td>
<td>8.9</td>
<td>6.8</td>
<td>0.5</td>
<td>6.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>9.2</td>
<td>10.1</td>
<td>8.6</td>
<td>3.8</td>
<td>7.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg</td>
<td></td>
<td>9.92</td>
<td>10.82</td>
<td>8.87</td>
<td>3.30</td>
<td>8.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-pregnant</td>
<td>1</td>
<td>13.3</td>
<td>13.3</td>
<td>12.9</td>
<td>11.1</td>
<td>12.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10.7</td>
<td>10.8</td>
<td>10.7</td>
<td>9.3</td>
<td>10.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>12.5</td>
<td>12.7</td>
<td>12.0</td>
<td>10.1</td>
<td>11.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8.4</td>
<td>8.7</td>
<td>8.1</td>
<td>5.7</td>
<td>7.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9.4</td>
<td>9.6</td>
<td>8.0</td>
<td>3.8</td>
<td>7.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>11.3</td>
<td>11.7</td>
<td>10.0</td>
<td>8.5</td>
<td>10.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg</td>
<td></td>
<td>10.93</td>
<td>11.13</td>
<td>10.28</td>
<td>8.08</td>
<td>10.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period and overall</td>
<td></td>
<td>Avg: 10.42</td>
<td>10.98</td>
<td>9.58</td>
<td>5.69</td>
<td>9.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

that may minimize the urgency of such examination for many samples of data. For a repeated effect (time or interaction), they recommend comparison of the F-ratio to a "conservative" tabular value of F, located after dividing the degrees of freedom for numerator and for denominator of the F-ratio by one less than the number of periods involved.

For the present example, the "conservative" F for testing either periods or interaction is $F(0.05, 1, 10)=4.96$, rather than the "usual" $F(0.05, 3, 30)=2.92$. In this case the true effects of time and interaction apparently are so large as to be readily evident even when the "conservative" F is used. Therefore, one need not examine the variance-covariance structure for these tests. If an F-ratio fails to exceed the "usual" tabular F, again there is no need to examine the structure, because the null hypothesis must be accepted in either case. It is only when an F-ratio falls between the "conservative" and "usual" tabular values of F that more rigorous procedures are indicated. These procedures will be discussed although they are not required for the data in table 1.

TABLE 2. ANALYSIS OF LITTER WEIGHT GAINS AS A COMPLETELY RANDOMIZED EXPERIMENT

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sums of squares</th>
<th>Mean squares</th>
<th>F-ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>1</td>
<td>42.56</td>
<td>42.56</td>
<td>9.17**</td>
</tr>
<tr>
<td>Periods</td>
<td>3</td>
<td>205.15</td>
<td>68.38</td>
<td>14.74**</td>
</tr>
<tr>
<td>T x P</td>
<td>3</td>
<td>33.30</td>
<td>11.11</td>
<td>2.56</td>
</tr>
<tr>
<td>Error</td>
<td>40</td>
<td>185.60</td>
<td>4.64</td>
<td>.......</td>
</tr>
</tbody>
</table>

** p<.01.

Examination of the Variance-Covariance Structure. In a repeat-measurement experiment with split-plot structure, the variance of data from each period and covariance of data from each possible pair of periods should be calculated within each treatment group. With four time-groups, there are four variances and six unique covariances (although 12 covariances can be written in a variance-covariance matrix because of symmetry). Also, one will need a "pooled" variance-covariance matrix ($S_p$), calculated by averaging common variance or covariance elements over all treatment groups, and a "uniform" variance-covariance matrix ($S_u$), calculated from $S_p$ by averaging all variances and all covariances (Appendix, A). Variance-covariance matrices for the two treatment groups and for the entire experiment are shown in table 4, with variances underlined. The determinants of these matrices are required and can be calculated by high-speed computer, or by desk calculator (Searle, 1966) if the matrices are small as in this example. For this example, the determinants are: det. $S_1=0.0086$, det. $S_2=0.0022$, det. $S_p=0.0188$ and det. $S_u=3.6551$.

The hypothesis that the true variance-covariance structure is identical from treatment to treatment should be examined first. Box (1949) has shown that the Chi-square distribution is an adequate approximation for testing this hypothesis if the number of treatment groups and periods is not large (say, no more than four or five each), and if adequate information is available to support each individual sample variance or covariance (say, 20...
## Table 3. Analysis of Litter Weight Gains as a Split Plot

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sums of squares</th>
<th>Mean squares</th>
<th>F-ratios</th>
<th>Expected mean squares *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between rats</td>
<td>(11)</td>
<td>42.56</td>
<td>42.56</td>
<td>2.54</td>
<td>$\sigma^2 + 4(1 + 3p)\sigma^2 + 24K^2 \rho$</td>
</tr>
<tr>
<td>Treatments</td>
<td>1</td>
<td>167.54</td>
<td>16.75</td>
<td>.....</td>
<td>$\sigma^2 + 4(1 + 3p)\sigma^2$</td>
</tr>
<tr>
<td>Rats within</td>
<td>10</td>
<td>167.54</td>
<td>16.75</td>
<td>113.59**</td>
<td>$\sigma^2 + (1 - \rho)\sigma^2 + 12K^2 \rho$</td>
</tr>
<tr>
<td>Within rats</td>
<td>(36)</td>
<td>205.15</td>
<td>68.38</td>
<td>19.72**</td>
<td>$\sigma^2 + (1 - \rho)\sigma^2 + 6K^2 \rho$</td>
</tr>
</tbody>
</table>

** P<.01, * * P<.001.

*a* represents random variation among samples at one time on one lactating rat (not replicated here); *K* terms are fixed elements of variation caused by treatments, periods, and interaction; *p* represents correlation between data from any two periods on the same lactating rat.

or more units per treatment group). A more accurate F-approximation (Box, 1950) should be used when information is limited, as in table 1 which has only six rats per group. Computation of the F-test is given in section B of the Appendix. The hypothesis of identical variance-covariance structure for pregnant and non-pregnant rats is acceptable for the data in table 3 ($F=0.84 < F_{0.05, 10, 478}=1.85$). If this hypothesis had been rejected, one should pursue some other procedures for testing treatment and period effects. For example, the non-repeated factor (treatments) could be tested within each period with an approximate F-test (Box, 1954), and the repeated factor (periods) could be tested *within* each treatment group by ordinary F-tests, if the variance-covariance structure is uniform within a group, or by the multivariate $T^2$ (Hotelling, 1931) if the structure is heterogeneous (Appendix, C, D). Or, use $T^2$ overall with approximate degrees of freedom (Yao, 1965).

Given that the true variance-covariance structure ostensibly is the same for each treatment group, one may pose the hypothesis that the true overall structure is uniform, i.e., like $S_o$. Box (1950) illustrated procedures for testing by Chi-Square or F-approximations (Appendix, B). The structure of the data in table 1 is highly heterogeneous ($F=5.41 > F_{0.01, 8, 1359}=2.55$). When the null hypothesis of uniformity is acceptable, one can rely on the usual F-tests of factorial effects in a split-plot (table 3). When it is not acceptable, the multivariate $T^2$ statistic is a logical alternative procedure for testing effects of periods, and treatment effects should be tested within each period by the approximate F-test (Box, 1954). Morrison (1967) and Cole and Grizzle (1966) discussed multivariate procedures, and Koch (1970) developed a non-parametric procedure for this kind of problem.

### Table 4. Variance-Covariance Matrices (Among Periods, $P_1$) for Pregnant ($S_1$), Non-Pregnant ($S_2$) and All Rats ($S_o$), and for Uniform Structure ($S_o$); Inverse of $S_o=(S_o^-)$

<table>
<thead>
<tr>
<th></th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$P_4$</th>
<th>$P_5$</th>
<th>$P_6$</th>
<th>$P_7$</th>
<th>$P_8$</th>
<th>$P_9$</th>
<th>$P_{10}$</th>
</tr>
</thead>
</table>

The **Multivariate $T^2$-Test.** The multivariate test (Hotelling, 1931) is proper for testing the repeated factor (periods) regardless of the form of the variance-covariance matrix, but Morrison (1967) has shown that it is less sensitive than the usual univariate F-test under conditions when both are valid (uniform variance-covariance), unless the true correlation between data from two periods exceeds 0.5 in absolute value. Therefore, one should use the multivariate procedure only when the F-test is invalid because of heterogeneous variance-covariance, or when the correlation between periods is large. The value of the multivariate procedure may be enhanced...
in some cases by using a priori estimates of variances and covariances determined from large quantities of similar data.

Computations for the $T^2$-test of period differences are given in section C of the Appendix. The inverse matrix, $S_{\text{rep}}^{-1}$, of matrix $S_{\text{rep}}$ (table 4) must be calculated by high-speed computer, or by desk calculator (Searle, 1966) if the matrix is as small as the one in this example. For the data in table 1, period differences were highly significant ($T^2 = 715.3 > T^2_{0.01} = 38.55$) as seen previously with the "conservative" F-test. Comparison of the "usual" test ($F/F_C = 21.1$), the "conservative" test ($F/F_C = 11.4$) and the $T^2$-test ($T^2/T^2_{0.01} = 18.6$) at the 99% significance level illustrates that the $T^2$-test is intermediate in probability to the two versions of the F-test.

A general test of interaction of treatment and period effects in a direct manner is not available. However, when only two treatment groups are used, as in this example, a more stringent null hypothesis may be of interest: that the profiles of response over time are identical for the two treatment groups (figure 1). The alternative to this hypothesis is that the two profiles are not coincident, whereas the alternative to a hypothesis of no interaction is merely that the two profiles are not parallel. Non-coincidence could reflect an average difference between the two treatments, interaction with periods or both. Morrison (1967, 1970) has discussed the test of coincidence, and Eaton (1969) generalized the procedure for more than two profiles. Computations for the test of identical profiles for the data in table 1 are given in section C of the Appendix. The profile of litter weight gains was significantly different for non-pregnant and pregnant lactating rats ($T^2 = 159.6 > T^2_{0.01} = 44.8$). It should be noted that this test leads to a different conclusion than the overall F-test of mean treatment differences (table 3) which was based on an invalid assumption of homogeneous variance-covariance structure, but is consistent with a significant difference in period 4 found by an approximate F-test (Appendix, D), and with the significant interaction of treatment and period (table 3) which is often of more interest than the main effect of treatments in this class of experiments.

**Summary**

An example of non-random repeated information of lactating rats is analyzed to clarify statistical procedures that avoid misleading inferences from such data. Recognition of split-plot structure and examination of heterogeneous correlation of data are emphasized. Methods are outlined for testing the variance-covariance structure, and multivariate procedures are discussed for testing hypotheses about effects of repeated factors (e.g., time) and for comparing profiles of responses (over time) for two treatment groups.

**Appendix**

Generalize the repeat-measurement split-plot experiment by considering data $y_{ijk}$ from treatment groups ($i=1---t$) each with animals ($j=1---n$), with non-random observations on each animal for several periods of time ($k=1---p$).

**A. Variance-Covariance Calculations**

Compute elements of a matrix $S_i$ for each treatment group $i$:

$$S_{ik} = \left[ \frac{\sum_j y_{ijk}^2 - (\sum_j y_{ijk})^2}{n} \right] / (n-1)$$

= Sample variance at period $k$ among $n$
REPEATED MEASUREMENTS OF ANIMALS

animals assigned to the same treatment group.

\[ S_{kk'} = \frac{\sum_{j=1}^{n} (y_{ijk}y_{ijk'}) - (\sum_{j=1}^{n} y_{ijk})}{n(n-1)} \]

= Sample covariance between periods \( k \) and \( k' \) for \( n \) animals assigned to the same treatment group.

Compute the "pooled" matrix, \( S_p = \sum_{i=1}^{t} S_i \); i.e., any element in \( S_p \) is the average of elements in the same position in the matrices for each treatment group. The "uniform" matrix, \( S_o \), contains variances all having the same value \( \sum_{k=1}^{p} S_{kk'}/p \) calculated from \( S_p \), and covariances all having the same value \( \sum_{k=1}^{p} \sum_{k'=2}^{p} S_{kk'}/(p(p-1)/2) \), i.e., the average of the \( C(p,2) \) unique covariances in \( S_p \), where \( C(p,2) \) is the number of combinations of \( p \) periods taken two at a time.

B. Tests of Hypotheses about Variance-Covariance Structure

For the null hypothesis of identical true structure for each treatment group, calculate:

\[ Z_1 = \left( \frac{(t+1)(2p^2+3p-1)}{6t(n-1)} \right) - \left( \frac{2+1}{6(2)(6-1)(4+1)} \right) = 0.43 \] for this example.

\[ Z_2 = (n-1) \left( \frac{\text{log e (det. } S_p \text{)}}{\text{log e (det. } S_t)} \right) \]

\[ = (6-1) \left( \frac{\text{log e (0.0188)}}{\text{log e (0.0086) + log e (0.0022)}} \right) \]

\[ = 0.1736 \text{ for this example.} \]

\[ Z_3 = \left( \frac{(p^2+p-4)}{6t^2(n-1)(p-1)} \right) = 0.21 \]

For the null hypothesis of true uniform structure (over all treatment groups), calculate:

\[ K_1 = \left( \frac{p(p+1)^2(2p-3)}{6t(n-1)(p-1)(p^2+p-4)} \right) \]

\[ = 0.1736 \text{ for this example.} \]

\[ K_2 = (n-1) \text{ log e (det. } S_p \text{/det. } S_o \text{)} \]

\[ = -2(6-1) \text{ log e (0.0188/3.6551)} \]

\[ = 52.73 \text{ for this example.} \]

\[ K_3 = \left( \frac{(p-1)(p+1)}{6t^2(n-1)(p-1)(p^2+p-4)} \right) = 0.375 \]

\[ f_1 = \left( \frac{(p^2+p-4)}{2} \right) \]

\[ f_2 = \left( \frac{f_1+2}{K_3-K_2} \right) = 38.55 \]

Reject the null hypothesis of uniform structure, i.e., consider the true variance-covariance structure heterogeneous.

C. Multivariate Tests of Hypotheses

For the null hypothesis of no real time effects (period differences), calculate:

\[ T_2 = tn[P'S^{-1}p] \]

\[ P'= \left( \frac{\bar{y}...-\bar{y} \bar{y}...2-\bar{y}}{\bar{y}...-\bar{y}} \right) \]

\[ = \left( \frac{10.42-9.17}{9.58-9.17} \right) \]

\[ = 59.609 \text{ for this example.} \]

Accept the null hypothesis of identical structure.

For the null hypothesis of true uniform structure (over all treatment groups), calculate:

\[ F = \left( \frac{K_2}{K_1} \right) - \frac{1}{(p+1)^2} \]

\[ = 5.41 > F_{0.1, 8, 1359} = 2.55 \]

Reject the null hypothesis of uniform structure, i.e., consider the true variance-covariance structure heterogeneous.
Reject the null hypothesis, and conclude that
an average time trend exists.

For the null hypothesis of identical profiles
of true response over time, calculate:

\[ T^2 = \left[ n_1 n_2 - (n_1 + n_2) \right] D[S^{-1} D] \]

where

\[ D'[\begin{array}{c}
\{ \bar{y}_{1.1} - \bar{y}_{2.1} \} \\
\{ \bar{y}_{1.2} - \bar{y}_{2.2} \}
\end{array}] = \begin{bmatrix}
\{ (9.92 - 10.93) (10.82 - 11.13) \\
(8.87 - 10.28) (3.30 - 8.08)
\end{bmatrix}
\]

is a row vector of mean differences
between the treatments for each period (table 1).

\[ (D'S^{-1} D) = \begin{bmatrix}
-48.253 & 36.900 & 26.179 \\
-11.049 & 53.198
\end{bmatrix} \]

\[ T^2 = \frac{(6)(6)/(6+6)(53.198)}{159.6} \]

\[ F_{0.01, p, n_1+n_2-p-1} = \frac{(6-6-2)4/(6+6-4-1)}{(5.714)(7.85)} = 44.85 \]

Reject the null hypothesis, and conclude that
the difference in sample profiles (figure 1) is
sufficient to indicate existence of a true differ-
ence in the populations of pregnant and non-
pregnant lactating rats.

**D. Approximate F-Test of Treatment Effects**

This test is appropriate when the variance-
covariance structure is heterogeneous (signifi-
cantly different for the various treatment
groups or non-uniform overall). Test treat-
ment effects within each period, as follows:

For example, within period 1,

\[ F = \left[ \sum \frac{y_{i.1}^2}{n_i} - (n-1) \frac{(\sum y_{i.1})^2}{n} \right] / \sum (S^2_i) \]

where \( y_{i.1} \) is a total for \( n \) animals in treatment
group \( i \), during period 1, and \( S^2_i \) is the average of sample variances in period 1 for the \( t \) different groups.

(Take this from \( S_0 \), table 4).

Compare with \( F_{0.06, t, f_2} \)

where \( f_2 = (n-1)(\sum S^2_i)^2 / \sum ((S^2_i)^2) \)

with \( S^2_i \) from the various treat-
ments during the period being considered.

Consider period 4 of the example:

\[ F = \left[ \frac{(19.8)^2 + (48.5)^2)}{6} - \frac{(19.8 + 48.5)^2}{12} \right] / (2-1) \]

\[ = (68.64/7.901) - 8.69 \]

\[ f_2 = \frac{5(8.032 + 7.770)^2}{(8.032)^2 + (7.770)^2} = 9.998 \] (in this case, essentially the same as unad-
justed, table 3).

\[ F_{0.01, 1, 16} = 4.96 \]

One may conclude that the treatment dif-
ference in period 4 is significant, i.e., the litter
weight gains for pregnant lactating rats are
less than for non-pregnant animals (figure 1).

Similar tests for the other three periods are
not significant.

**Literature Cited**

Box, G. E. P. 1949. A general distribution theory for
a class of likelihood criteria. Biometrika 36:317.

Box, G. E. P. 1950. Problems in the analysis of

Box, G. E. P. 1954. Some theorems on quadratic
forms applied in the study of analysis of variance
problems. II. Effects of inequality of variance and
of correlation of errors in the two-way classifica-

Cochran, W. G. and G. Cox, 1957. Experimental De-
signs (2nd Ed.). John Wiley and Sons, Inc., New
York.

Cole, J. W. L. and J. E. Grizzle. 1966. Applications
of multivariate analysis of variance to repeated

Ass. 64:1318.

of Box's results on the use of the F distribution in

Hotelling, H. 1931. The generalization of Student's ratio.

Koch, G. G. 1970. The use of non-parametric methods
in the statistical analysis of a complex split plot

Morrison, D. F. 1967. Multivariate Statistical Meth-

Morrison, D. F. 1970. The optimal spacing of re-

nucleic acid, hydroxyproline, and hexosamine
of pregnant rats during lactation and post-lactational

Searle, S. 1966. Matrix Algebra for the Biological

Methods. (6th Ed.). The Iowa State University
Press, Ames.

Freeman and Co., San Francisco.

Yao, Ying. 1965. An approximate degrees of freedom
solution to the multivariate Behrens-Fisher prob-
lem. Biometrika 52:139.