

UNCLASSIFIED

Defense Technical Information Center
Compilation Part Notice

ADP014608

TITLE: Analysis of a Function in Collaborative Experimentation

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Proceedings of the Eighth Conference on the Design of
Experiments in Army Research Development and Testing

To order the complete compilation report, use: ADA419759

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:

ADP014598 thru ADP014630

UNCLASSIFIED

ANALYSIS OF A FUNCTION IN COLLABORATIVE EXPERIMENTATION

Walter D. Foster
Biomathematics Division
Fort Detrick, Frederick, Maryland

I. INTRODUCTION. The usual objective in collaborative or referee experimentation is to make comparisons among the set of participants with the over-all criterion that stations be no more diverse than runs at a station. Thus, station means and variances are the values for interstation comparison. This paper is concerned with the response variable for a particular class of referee experimentation and its analysis.

In this collaborative experiment, each of five laboratories ran a series of aerosol tests in which P. tularensis tagged with radioactive phosphorous (P^{32}) was aerosolized in rotating drums and sampled at eight points in time over a 22-hour period. The five laboratories with identical equipment achieved the series of tests at approximately the same time, going to extreme lengths to achieve homogenous methodology. Three treatments were introduced consisting of three relative humidity conditions in the rotating drums of 20%, 50% and 80%. Two aerosols or runs were completed per humidity at each participating laboratory on a randomized basis. It is of interest to note that three separate nations were represented in these five stations.

It was the objective of this experimentation to

- (1) Compare station means,
- (2) Compare station variances,
- (3) To identify stations whose results did not conform to those of the others,
- (4) To examine the station by treatment interaction, i.e., whether the differences between treatments were consistent from one station to another.

II. DEFINITION OF THE RESPONSE VARIABLE. When an aerosol is monitored over a period of time, the measurement usually taken is the concentration at a series of points in time. Thus, the definition of the response variable to be analyzed could be the concentration,

given a particular set of sampling times. However, this concept is likely to ignore the design restriction that only runs are random, not sampling points in a run. A second and better response variable is the function describing concentration and its change in time. Such a function in aerobiology is called a decay function. Previous research has identified a reasonably simple expression which is excellent for summarizing the course of an aerosol in time:

$$C = C_0 (t + 1)^{-b} e^{-kt}$$

The usual univariate approach to the analysis of a function such as the one given above would be to analyze separately the parameters of this function, C_0 , b , k . However, not only are these parameters known to be correlated because of the design of the experiment but they are also known to be stochastically correlated from one aerosol run to another. Therefore, it is the purpose here to show how the entire decay function, identified as the response variable, can be analyzed and interpreted through the usual analysis of variance technique.

III. ANALYSIS OF VARIANCE OF THE DECAY FUNCTION. With the decay function as the response variable, the following analysis of variance has been accorded this response for the purpose of examining stations levels, variability, and station by treatment interaction. The complete analysis of variance is shown in Table I in detailed form where all of the objectives have been answered. Its construction is given in a separate section.

TABLE I.

A. V. OF DECAY FUNCTION FOR STATIONS AND TREATMENTS

<u>Line</u>	<u>Source</u>	<u>df</u>	<u>MS</u>
15	Mean	3	387.1591
16	Stations	12	.1737
17	A vs Rest	3	.5894
18	Among Rest	9	.0352
19	Treatments	6	.0409
20	S x T	24	.0151
21	Runs in S x T	45	.0195
22	Runs in 20%	15	.0315
23	Runs in 50%	15	.0129
24	Runs in 80%	15	.0118
25	Deviations	150	.0014
26	TOTAL		

The following brief interpretation is accorded the analysis of variance shown in Table I in order to provide specific answers to the objectives of this experiment. Reading from the bottom of the table, the runs have been pooled over stations per treatment affording a test of homogeneity of variance from one treatment to another in lines 22-24. This departure from the original objective is better achieved than the original for estimating station variability because of the limited number of runs per treatment. There is a suggestion that the runs were less homogeneous at the 20% humidity than at the other two. In line 20, it is clear that the station by humidity interaction, if not zero, was small. On the other hand in line 16, differences among stations were obviously large compared to runs in S x T, line 21. The contrast of A versus the remaining stations, line 17, accounted for a large proportion of the station variability, with the variation attributed to the remaining stations being scarcely larger than the variation among trials at a given station. The purpose of this partition in line 17 was to investigate whether the variation among the remaining stations has been reduced to magnitude of trial-to-trial variation. Further partition is in order so long as it could be helpful in identifying and possibly eliminating factors at stations causing station departures.

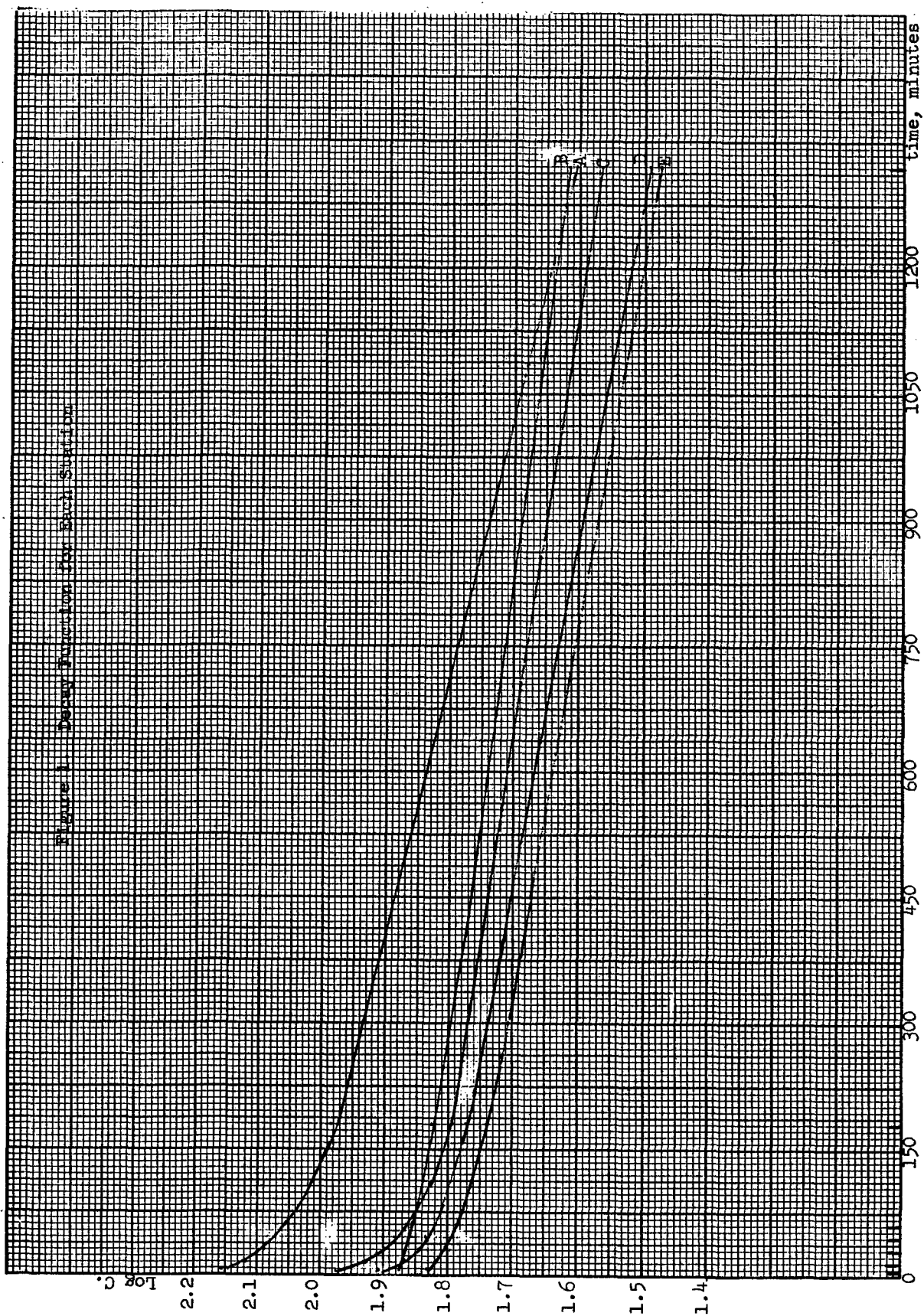
This brief interpretation was developed completely on the basis of the analysis of variance in Table I. It would be desirable to present a tables of means to accompany the variance analysis. This is the point at which multivariate techniques in general are at a disadvantage, for there is no plainly defined quantity which is easily tabled. Two suggestions are given here as a means by which the interpretation can be visualized; these are first by graphs and secondly by the coefficients of the decay function. The graphs for each station are given in Figure 1 where the values have been averaged over all three humidity conditions. The coefficients computed as estimates of the parameters of the decay function are given below.

Values of Decay Function Constants

	Stations				
	A	B	C	D	E
Log C_0	2.227	2.053	1.879	1.868	1.972
$b \times 10^1$	.868	.994	.053	.406	.839
$k \times 10^3$	.264	.193	.219	.205	.073

It is appreciated that neither of these means of visualizing station differences is perfect; nevertheless, they are suggested here as the best which are easily available.

A few remarks are necessary here before describing in the next section the technique for the analysis of variance of a decay function. The question of auto-correlation always seems to appear in problems in time series such as these. However, it is contended here that because of the function approach the question of auto-correlation of the successive C_i does not arise. Only the residuals are important, and when the decay function is found to provide an excellent summary of the change of concentration in time the residuals may be considered as mutually independent. A second remark has to do with the potential use of the results of this kind of referee experimentation. With the variation noted here and appraised to be acceptable, these data afford a basis for constructing a quality control approach to future aerosol runs in which aberrant points, runs, and even stations may be readily identified.



IV. CONSTRUCTION OF THE ANALYSIS OF VARIANCE OF A DECAY FUNCTION. On an individual run basis, the familiar partition of variation is obtained as shown in Table II with the exception that no correction is shown separately for the mean -- the p parameters of the decay function are shown together.

TABLE II.

A. V. OF DECAY FUNCTION FOR A SINGLE RUN

<u>Line</u>	<u>Source</u>	<u>df</u>	<u>e. g. df</u>
1	Function	p	3
2	<u>Deviations</u>	$\frac{n-p}{n}$	$\frac{5}{8}$
3	TOTAL		

The second step is to compute the analysis of variance for each station over the r runs for a given treatment as is shown in Table III. The sum of squares for line 4 are obtained as usual where the function is fitted to the entire set of values for the r runs and the computation is achieved on a per item (or per value) basis. The sum of squares for line 5 is obtained easily merely by summing the sum of squares for line 1 in Table II for the various runs and subtracting line 4. Similarly, line 6, deviations in runs, is obtained by summing the values in line 2 over all runs.

TABLE III.

A. V. OF DECAY FUNCTION FOR A STATION AND A TREATMENT

<u>Line</u>	<u>Source</u>	<u>df</u>	<u>e. g. df</u>
4	Mean	p	3
5	Among runs	$p(r-1)$	3
6	<u>Deviations in runs</u>	$\frac{r(n-p)}{rn}$	$\frac{10}{16}$
7	TOTAL		

A small digression may be helpful at this point to explain the degrees of freedom shown thus far in the analysis of variance. The degrees of freedom in line 5 are shown to be the usual degrees of freedom for runs, $r-1$, multiplied by the number of parameters to be estimated in the decay function. Although these parameters are known not to be independent, they continue to be identified as restrictions in the least squares process for estimation and as such must be deducted as degrees of freedom. It is not likely that a further partition of these degrees of freedom could be achieved in a manner such as to show contrasts among the parameters themselves.

With the introduction of t treatments at a station, the analysis of variance as outlined in Table IV is appropriate for each station, where the partition is basically a nested one. As before, the function is fitted over all points in order to provide the sum of squares due to the function, line 8. The sum of squares for treatments is obtained through a two step procedure. First, the sums of squares shown in line 4 of Table III for each treatment are added. Then the sum of squares for the mean in line 8 is subtracted, the difference being specifically that due to variation among treatments and is entered in line 9. The sum of squares for runs in treatments, line 10, is obtained by summing the sums of squares for each trial separately for that particular treatment, i. e., the sum of lines 5 for that station. They can also be listed in partition as in lines 11 and 12 of Table IV. Similarly, the sum of squares for deviations are obtained by pooling for line 13.

TABLE IV.

A. V. OF DECAY FUNCTION AT STATION A WITH TREATMENTS

<u>Line</u>	<u>Source</u>	<u>df</u>	<u>e. g. df</u>
8	Mean	p	3
9	Treatments	$p(t-1)$	6
10	Runs in T	$pt(r-1)$	9
11	in T_1	$p(r-1)$	3
12	in T_2	$p(r-1)$	3
	etc	etc	3
13	<u>Deviations</u>	<u>$rt(n-p)$</u>	<u>30</u>
14	TOTAL	trn	48

The construction of the over-all analysis of variance as shown in Table I continues to be based upon the previous tables in a sort of a building block arrangement. The mean, line 15, is obtained by finding the sum of squares due to the function when fitted to all of the points in the combined collaborative experiment. Line 16 is obtained by a two step procedure: the sum of squares for line 8 in Table IV is summed over the s stations; from this sum of lines 8 the sum of squares in line 15 is subtracted. The difference then represents the sum of squares due to stations averaged over treatments.

The partition of the station sum of squares as initiated in line 17 depends upon which station appears to show the greatest departure from the other stations, following the philosophy given briefly in the interpretation of the example above. Assuming that this identification of the greatest departure can be made from a study of the graphs, line 17 then represents the contrast between the station with the maximum departure and the rest of the stations. This partition is accomplished in a three step procedure as follows. The sums of squares given in line 8 of Table IV are added for the four stations marked as "rest". This sum is entered as line "a" in the ancillary computation table below. The second step is to compute the sum of squares for the function when fitted to all the points represented by the four stations combined as "rest", having excluded the station with the maximum departure from the computation--line "b" below. The third step is to subtract the sum of squares in line "b" from the sum of squares in line "a", giving the "among rest" sum of squares as shown in line "c". Finally, the subtraction of line "c" sum of squares from line 16 is entered in line 17 and is identified as the contrast station A versus "rest". Further orthogonal partitioning for other "departures" can be computed in this fashion.

Ancillary Computation for Table I

<u>Line</u>	<u>Source</u>
a	Sum of line 8 for "rest" stations
b	Mean for "rest"
c	$a - b =$ among "rest" stations

A new computation is required for line 19, the sum of squares due to treatments. This is accomplished by considering all points for the first treatment including those for the various stations and fitting the decay

function. This is achieved for each treatment. These sum of squares are added over the various treatments. From this over-all sum, the value in line 15 is subtracted, giving the variation among treatments averaged over stations.

The interaction term, station by treatment, as shown in line 20, is obtained in the usual way. Briefly, it consists of summing line 4 over all stations and treatments. From this sum are subtracted lines 15, 16 and 19.

Line 21 is obtained easily by summing all lines in Table III. The partition of line 21 as shown in lines 22 and 23 is easily accomplished according to the purpose at hand merely by restricting the summing to the category desired.

Missing values will complicate this analysis and indeed will render the partition non-orthogonal if missing values are not restored to the analysis. Therefore, it is recommended that a simple procedure for estimating these missing values such as computing the value according to the function as estimated from the remainder of the points being inserted with one degree of freedom per missing value being subtracted from the degree of freedom assigned to deviations. Note that in the simpler analyses which are completely nested orthogonality does not depend upon equal numbers.