

Multiple Factor Analysis of Variance

KEY WORDS air pollution, dioxin, furan, incineration, samplers, ANOVA, analysis of variance, factorial experiment, sampling error.

Environmental monitoring is expensive and complicated. Many factors may contribute variation to measured values. An obvious source of variation is the sampling method. An important question is: "Do two samplers give the same result?" This question may arise because a new sampler has come on the market, or because a monitoring program needs to be expanded and there are not enough samplers of one kind available.

It might seem natural to compare the two (or more) available sampling methods under a fixed set of conditions. This kind of experiment would estimate random error under only that specific combination of conditions. The samplers, however, will be used under a variety of conditions. A sampler that is effective under one condition may be weak under others. The error of one or both samplers might depend on plant operation, weather, concentration level being measured, or other factors. The variance due to laboratory measurements may be a significant part of the total variance. Interactions between sampling methods and other possible sources of variation should be checked. The experimental design should take into account all these factors.

Comparing two samplers under fixed conditions pursues the wrong goal. A better plan would be to assess performance under a variety of conditions. It is important to learn whether variation between samplers is large or small in comparison with variation due to laboratory analysis, operating conditions, etc. A good experiment would provide an analysis of variance of all factors that might be important in planning a sampling program.

It is incorrect to imagine that one data point provides one piece of information and therefore the information content of a data set is determined entirely by the number of measurements. The amount of information available from a fixed number of measurements increases dramatically if each observation contributes to estimating more than one parameter (mean, factor effect, variance, etc.). An exciting application of statistical experimental design is to make each observation do double duty or even triple or heavier duty. However, any valid statistical analysis can only extract the information existing in the data at hand. This content is largely determined by the experimental design and cannot be altered by the statistical analysis. This chapter discusses an experimental design that was used to efficiently evaluate four factors that were expected to be important in an air quality monitoring program. The experiment is based on a factorial design (but not the two-level design discussed in Chapter 27). The method of computing the results is not discussed because this can be done by commercial computer programs. Instead, discussion focuses on how the *four-factor analysis of variance* is interpreted. References are given for the reader who wishes to know how such experiments are designed and how the calculations are done (Scheffe, 1959).

Case Study: Sampling Dioxin and Furan Emissions from an Incinerator

Emission of dioxins and furans from waste incinerators has been under investigation in many countries. It is important to learn whether different samplers (perhaps used at different incinerators or in different cities or countries) affect the amount of dioxin or furan measured. It is also important to assess whether differences, if any, are independent of other factors (such as incinerator loading rate and feed materials which change from one sampling period to another).

TABLE 26.1
Dioxin and Furan Data from a Designed Factorial Experiment

Sample Period Sampler	1		2		3		4	
	A	B	A	B	A	B	A	B
Dioxins								
Sum TetraCDD	0.4	1.9	0.5	1.7	0.3	0.7	1.0	2.0
Sum PentaCDD	1.8	28	3.0	7.3	2.7	5.5	7.0	11
Sum HexaCDD	2.5	24	2.6	7.3	3.8	5.1	4.7	6.0
Sum HeptaCDD	17	155	16	62	29	45	30	40
Sum OctoCDD	7.4	55	7.3	28	14	21	12	17
Furans								
Sum TetraCDF	4.9	26	7.8	18	5.8	9.0	13	13
Sum PentaCDF	4.2	31	11	22	7.0	12	17	24
Sum HexaCDF	3.5	31	11	28	8.0	14	18	19
Sum HeptaCDF	9.1	103	32	80	32	41	47	62
Sum OctoCDF	3.8	19	6.4	18	6.6	7.0	6.7	6.7

Note: Values shown are concentrations in ng/m^3 normal dry gas at actual CO_2 percentage.

The data in Table 26.1 were collected at a municipal incinerator by the Danish Environmental Agency (Paltessen, 1987). Two different kinds of samplers were used to take simultaneous samples during four 3.5-hour sampling periods, spread over a three-day period. Operating load, temperature, pressure, etc. were variable. Each sample was analyzed for five dioxin groups (TetraCDD, PentaCDD, HexaCDD, HeptaCDD, and OctoCDD) and five furan groups (TetraCDF, PentaCDF, HexaCDF, HeptaCDF, and OctoCDF). The species within each group are chlorinated to different degrees (4-, 5-, 6-, 7, and 8 chlorine atoms per molecule). All analyses were done in one laboratory.

There are four factors being evaluated in this experiment: two kinds of samplers (S), four sampling periods (P), two dioxin and furan groups (DF), five levels of chlorination within each group (CL). This gives a total of $n = 2 \times 4 \times 2 \times 5 = 80$ measurements. The data set is completely balanced: all conditions were measured once with no repeats. If there are any missing values in an experiment of this kind, or if some conditions are measured more often than others, the analysis becomes more difficult (Milliken and Johnson, 1992).

When the experiment was designed, the two samplers were expected to perform similarly but that variation over sampling periods would be large. It was also expected that the levels of dioxins and furans, and the amounts of each chlorinated species, would be different. There was no prior expectation regarding interactions. A four-factor analysis of variance (ANOVA) was done to assess the importance of each factor and their interactions.

Method: Analysis of Variance

Analysis of variance addresses the problem of identifying which factors contribute significant amounts of variance to measurements. The general idea is to partition the total variation in the data and assign portions to each of the four factors studied in the experiment and to their interactions.

Total variance is measured by the total residual sum of squares:

$$\text{Total SS} = \sum_{\text{all obs}} (y_{\text{obs}} - \bar{y})^2$$

where the residuals are the deviations of each observation from the grand mean

$$\bar{y} = \frac{1}{n} \sum_{\text{all obs}} y_i$$

of the $n = 80$ observations. This is also called the total adjusted sum of squares (corrected for the mean). Each of the n observations provides one degree of freedom. One of them is consumed in computing the grand average, leaving $n - 1$ degrees of freedom available to assign to each of the factors that contribute variability. The total SS and its $n - 1$ degrees of freedom are separated into contributions from the factors controlled in the experimental design. For the dioxin/furan emissions experiment, these sums of squares (SS) are:

$$\text{Total SS} = \text{Periods SS} + \text{Samplers SS} + \text{Dioxin/Furan SS} + \text{Chlorination SS} \\ + \text{Interaction(s) SS} + \text{Error SS}$$

Another approach is to specify a general model to describe the data. It might be simple, such as:

$$y_{ijk} = \bar{y} + \alpha_i + \beta_j + \gamma_k + \lambda_{ij} + (\text{interaction terms}) + \epsilon_i$$

where the Greek letters indicate the true response due to the four factors and ϵ_i is the random residual error of the i th observation. The residual errors are assumed to be independent and normally distributed with mean zero and constant variance σ^2 (Rao, 1965; Box et al., 1978).

The assumptions of independence, normality, and constant variance are not equally important to the ANOVA. Scheffe (1959) states, "In practice, the statistical inferences based on the above model are not seriously invalidated by violation of the normality assumption, not... by violation of the assumption of equality of cell variances. However, there is no such comforting consideration concerning violation of the assumption of statistical independence, except for experiments in which randomization has been incorporated into the experimental procedure."

If measurements had been replicated, it would be possible to make a direct estimate of the error sum of squares (σ^2). In the absence of replication, the usual practice is to use the higher-order interactions as estimates of σ^2 . This is justified by assuming, for example, that the fourth-order interaction has no meaningful physical interpretation. It is also common that third-order interactions have no physical significance. If sums of squares of third-order interactions are of the same magnitude as the fourth-order interaction, they can be pooled to obtain an estimate of σ^2 that has more degrees of freedom.

Because no one is likely to manually do the computations for a four-factor analysis of variance, we assume that results are available from some commercial statistical software package. The analysis that follows emphasizes variance decomposition and interpretation rather than model specification.

The first requirement for using available statistical software is recognizing whether the problem to be solved is one-way ANOVA, two-way ANOVA, etc. This is determined by the number of factors that are considered. In the example problem there are four factors: S, P, DF, and CL. It is therefore a four-way ANOVA.

In practice, such a complex experiment would be designed in consultation with a statistician, in which case the method of data analysis is determined by the experimental design. The investigator will have no need to guess which method of analysis, or which computer program, will suit the data. As a corollary, we also recommend that happenstance data (data from unplanned experiments) should not be subjected to analysis of variance because, in such data sets, randomization will almost certainly have not been incorporated.

Dioxin Case Study Results

The ANOVA calculations were done on the natural logarithm of the concentrations because this transformation tended to strengthen the assumption of constant variance.

The results shown in Table 26.2 are the complete variance decomposition, specifying all sum of squares (SS) and degrees of freedom (df) for the main effects of the four factors and all interactions between the four factors. These are produced by any computer program capable of handling a four-way ANOVA

TABLE 26.2
Variance Decomposition of the Dioxin/Furan Incinerator Emission Data

Source of Variation	SS	df	MS	F
S	18.3423	1	18.3423	573
CL	54.5564	4	13.6391	426
DF	11.1309	1	11.1306	348
DF × CL	22.7618	4	5.6905	178
S × P	9.7071	3	3.2357	101
P	1.9847	3	0.6616	21
DF × P	1.1749	3	0.3916	12.2
DF × S	0.2408	1	0.2408	7.5
P × CL	1.4142	12	0.1179	3.7
DF × P × CL	0.8545	12	0.0712	2.2
S × P × CL	0.6229	12	0.0519	1.6
S × S × CL	0.0895	4	0.0224	0.7
DF × S × P × CL	0.0826	4	0.0206	0.6
DF × S × P	0.2305	12	0.0192	0.6
F	0.0112	3	0.0037	0.1

F calculated using $\sigma^2 = 0.032$, which is estimated with 27 degrees of freedom.

(e.g., SAS, 1982). The main effects and interactions are listed in descending order with respect to the mean sums of squares (MS = SS/df).

The individual terms in the sums of squares column measure the variability due to each factor plus some random measurement error. The expected contribution of variance due to random error is the random error variance (σ^2) multiplied by the degrees of freedom of the individual factor. If the true effect of the factor is small, its variance will be of the same magnitude as the random error variance. Whether this is the case is determined by comparing the individual variance contributions with σ^2 , which is estimated below.

There was no replication in the experiment so no independent estimate of σ^2 can be computed. Assuming that the high-order interactions reflect only random measurement error, we can take the fourth-order interaction, DF × S × P × CL, as an estimate of the error sum of squares, giving $\hat{\sigma}^2 = 0.2305/12 = 0.0192$. We note that several other interactions have mean squares of about the same magnitude as the DF × P × CL interaction and it is tempting to pool these. There are, however, no hard and fast rules about which terms may be pooled. It depends on the data analyst's concept of a model for the data. Pooling more and more degrees of freedom into the random error term will tend to make $\hat{\sigma}^2$ smaller. This carries risks of distorting the decision regarding significance and we will follow Pallesen (1987) who pooled only the fourth-order and two third-order interactions (S × P × CL and of S × P × DF) to estimate $\hat{\sigma}^2 = (0.2305 + 0.6229 + 0.0112)/(12 + 12 + 3) = 0.8646/27 = 0.032$.

The estimated error variance ($\hat{\sigma}^2 = 0.032 = 0.18'$) on the logarithmic scale can be interpreted as a measurement error with a standard deviation of about 18% in terms of the original concentration scale. The main effects of all four factors are all significant at the 0.05% level. The largest source of variation is due to differences between the two samplers. Clearly, it is not acceptable to consider the samplers as equivalent. Presumably sampler B gives higher concentrations (Table 26.1), implying greater efficiency of contaminant recovery. The differences between samplers is much greater than differences between sampling periods, although 'periods' represents a variety of operating conditions.

The interaction of the sampler with dioxin/furan groups (S × DF) was small, but statistically significant. The interpretation is that the difference between the samplers changes, depending on whether the contaminant is dioxin or furan. The S × P interaction is also significant, indicating that the difference between samplers was not constant over the four sampling periods.

The *a priori* expectation was that the dioxin and furan groups (DF) would have different levels and that the amounts of the various chlorinated species (CL) with chemical groups would not be equal. The large mean squares for DF and CL supports this.

Comments

When the experiment was planned, variation between sampling periods was expected to be large and differences between samplers were expected to be small. The data showed both expectations to be wrong. The major source of variation was between the two samplers. Variation between periods was small, although statistically significant.

Several interactions were statistically significant. These, however, have no particular practical importance until the matter of which sampler to use is settled. Presumably, after further research, one of the samplers will be accepted and the other rejected, or one will be modified. If one of the samplers were modified to make it perform more like the other, this analysis of variance would not represent the performance of the modified equipment.

Analysis of variance is a useful tool for breaking down the total variability of designed experiments into interpretable components. For well-designed (complete and fully balanced) experiments, this partitioning is unique and allows clear conclusions to be drawn from the data. If the design contains missing data, the partition of the variation is not unique and the interpretation depends on the number of missing values, their location in the table, and the relative magnitude of the variance components (Cohen and Cohen, 1983).

References

- Box, G. E. P., W. G. Hunter, and J. S. Humer (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. New York, Wiley Interscience.
- Cohen, J. and P. Cohen (1983). *Applied Multiple Regression & Correlation Analysis for the Behavioral Sciences*, 2nd ed., New York, Lawrence Erlbaum Assoc.
- Milliken, G. A. and D. E. Johnson (1992). *Analysis of Messy Data, Vol. I: Designed Experiments*, New York, Van Nostrand Reinhold.
- Milliken, G. A. and D. E. Johnson (1989). *Analysis of Messy Data, Vol. II: Nonreplicated Experiments*, New York, Van Nostrand Reinhold.
- Pallesen, L. (1987). "Statistical Assessment of PCDD and PCDF Emission Data," *Waste Manage. Res.*, 5, 367-379.
- Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*, New York, John Wiley.
- SAS Institute Inc. (1982). *SAS User's Guide: Statistics*, Cary, NC.
- Scheffé, H. (1959). *The Analysis of Variance*, New York, John Wiley.

Exercises

- 26.1 Dioxin and Furan Sampling. Reinterpret the Pallesen example in the text after pooling the higher-order interactions to estimate the error variance according to your own judgment.
- 26.2 Ammonia Analysis. The data below are the percent recovery of 2 mg/L of ammonia (as $\text{NH}_4\text{-N}$) added to wastewater final effluent and tap water. Is there any effect of pH before distillation or water type?

pH Before Distillation	Final Effluent (initial conc. = 13.8 mg/L)	Tap Water (initial conc. = 0.1 mg/L)
9.5*	98	100
6.0	100	88
6.5	102	99
7.0	98	99
7.5	105	103
8.0	102	101
	99	95
	99	95
	98	94

* Buffered.

Source: Dhaliwal, B. S., *J. WPCF*, 57, 1036-1039.