Research in Conservation

1988

1

Terry J. Reedy
Chandra L. Reedy

Statistical Analysis in Art Conservation Research

THE GETTY CONSERVATION INSTITUTE

Statistical Analysis in Art Conservation Research

Research in Conservation

1988

1

Terry J. Reedy
Chandra L. Reedy

Statistical Analysis in Art Conservation Research

THE GETTY CONSERVATION INSTITUTE

Printed in the United States of America.

Library of Congress Cataloging-in-Publication Data

Reedy, Terry J., 1947-
Statistical analysis in art conservation research.

(Research in conservation)
Bibliography: p.
Includes index.
1. Art-Conservation and restoration-Research-Statistical methods.
I. Reedy, Chandra L., 1953- II. Title. III. Series.
N8555.R44 1987 702'.8'8072 88-2994
ISBN 0-89236-097-6

The Getty Conservation Institute

The Getty Conservation Institute (GCI), an operating program of the J. Paul Getty Trust, was created in 1982 to enhance the quality of conservation practice in the world today. Based on the belief that the best approach to conservation is interdisciplinary, the Institute brings together the knowledge of conservators, scientists, and art historians. Through a combination of in-house activities and collaborative ventures with other organizations, the Institute plays a catalytic role that contributes substantially to the conservation of our cultural heritage. The Institute aims to further scientific research, to increase conservation training opportunities, and to strengthen communication among specialists.

Research in Conservation

This reference series is born from the concern and efforts of the Getty Conservation Institute to publish and make available the findings of research conducted by the GCI and its individual and institutional research partners, as well as state-of-the-art reviews of conservation literature. Each volume will cover a separate topic of current interest and concern to conservators. Publication will be on an irregular schedule, but it is expected that four to six volumes will be available each year. Annual subscriptions and individual titles are available from the GCI.

Contents

The Authors

Terry J. Reedy

Dr. T. Reedy has degrees in mathematics, operations research, and ecology. He has been a consulting statistician since 1979 in the Biomathematics Unit of the Center for Ulcer Research and Education, in the University of California, Los Angeles, Medical School. This has given him broad experience with the practical problems of data analysis in scientific research. He also works as an independent consultant and was introduced to the problems of statistics in art history and archaeometry while helping his wife Chandra with her Master's and Ph.D theses projects.

Chandra L. Reedy

Dr. C. Reedy received her Ph.D in archaeology from the University of California, Los Angeles, in 1986, where her areas of specialization were archaeometry and Himalayan art and archaeology. She is currently an Andrew W. Mellon Fellow in Conservation Research at the Los Angeles County Museum of Art. Her particular interest is introducing scientific methods into the study of art.

Preface

This technical report reviews the use of statistics in art conservation research. Our aim is to examine how statistical analyses have been handled in published conservation research studies and to suggest alternative approaches. All components of data analysis—including experimental design, data organization, and statistical techniques—are evaluated.

History

This report was produced as part of a contract between the Getty Conservation Institute Scientific Research Program, and the Los Angeles County Museum of Art, Conservation Center. The purpose of the contracted project was to explore the use of statistics in art conservation and archaeometry. The original version of this report was presented to the Getty Conservation Institute Scientific Research Program for the purpose of helping them with the use of statistics in their internal and external research projects. At their request we have rewritten it for a wider audience.

Production

The text was edited with the WordPerfect word processing program on MSDOS microcomputers, the AT&T 6300 and a generic AT-compatible. The statistical analyses were carried out with various programs from the BMDP Statistical Software package running on both a UNIX desktop computer and the MSDOS machines. Text was prepared using Xerox Ventura Desktop Publisher 1.1 on an IBM PC-AT and output on a Linotronic 300 at 1270 DPI.

Acknowledgments

Pieter Meyers, Head of Conservation, and John Twilley, Senior Research Chemist of the Conservation Center at the Los Angeles County Museum of Art, both read and commented on several versions of this report. It was also reviewed by Frank Preusser, James Druzik, Michele Derrick, Miguel Angel Corzo, and John Perkins, all from the Getty Conservation Institute. We thank all of these people for their helpful comments, which have measurably improved our presentation.

Chapter 1

Statistical Analysis and Art Conservation Research

Introduction

Statistics

The origin of the term *statistics* is related to "state" and "status." Numbers such as population and tax revenue, which are characteristics of a state or nation, are statistics in the classical sense. In modern usage, a statistic is any number calculated from raw research data. Some statistics, such as counts, means, and standard deviations, describe a population or sample. Other statistics, such as t and F statistics, are used to test hypotheses about the population.

More broadly, statistics is the art and science encompassing the theory and techniques developed for calculating and using such numbers. In the broadest sense, statistics is the application of the scientific method to data collection and analysis and the incorporation of rigorous data analysis into the scientific method.

Statistics are used to describe objects, estimate the characteristics of a population from a sample, and test hypotheses or ideas about the subject of a study. The latter two uses have in common the problem of making decisions in the face of uncertainty or variability. One of the characteristics of the statistical approach is to admit the existence of, measure, and make the best of imperfection, error, and variation.

Art Conservation

Many projects in the field of conservation research require statistical analysis to make optimal use of the data collected. The purpose of studies with numerical data is often to evaluate and compare conditions and treatments. Such comparisons are a classical statistical problem. (This goal is quite different from making a qualitative decision between alternative mechanisms or competing theories.) There are also methods, often newer and less well known outside of the statistical journals, for making sense out of categorical data collected for other reasons.

Art materials, and especially art objects, have two particular characteristics that must be taken into account in any statistical analysis. They are internally heterogeneous and individually distinct in composition, form, and history. This variability necessitates attentive consideration of the statistical procedures used at each stage of data analysis. Ideally, selection of the most appropriate method of statistical analysis for a given project is a result of careful reflection on both the scientific questions to be answered and the structure of the data collected.

Organization

We reviewed 320 papers published between 1980 and 1986 in four English-language conservation journals, which are abbreviated throughout this report as follows:

JC *Journal of the American Institute of Conservation*
SC *Studies in Conservation*
TB *National Gallery Technical Bulletin*
PP *AIC Preprints*

Details about the issues covered and the number of articles from each are given in Chapter 5.

The second section of this first chapter summarizes the major findings of our research, including recommendations for areas where improvements in statistical procedures are the most crucial. The chapters that follow present in detail the motivations and principles behind these recommendations.

There are three phases to an art conservation project:

1. determination of the composition of the art object or material
2. consideration of how it has or might deteriorate
3. application of conservation materials and methods to remedy current damage or prevent further damage

Each of these three phases differ somewhat in the types of research questions asked, the experimental methods used, and the statistical methods required. Most papers focus on just one phase. We therefore split the papers into three groups, one for each phase, and discuss each group in a separate chapter. Chapter 2 focuses on problems and methods specific to studies of art object composition. Chapter 3 does the same for studies of art material deterioration, and Chapter 4 for studies involving the effectiveness of conservation treatments and materials.

Chapter 5 presents a statistical analysis of the statistical methods used in the 320 papers reviewed. Several numerical scores related to the organization and use of various data analytical and statistical procedures, as well as identifying information, are tabulated for each paper. The resulting data table is then analyzed to answer several specific questions.

Presentation

The middle three chapters mix discussions of general principles of statistical analysis that are especially pertinent to conservation research with examples drawn from the literature to illustrate the application of those principles. Suggested alternatives and improvements are presented. In some cases, published data are reanalyzed to show the results that can be obtained by the proposed method of statistical analysis.

The purpose of reviewing published papers and using their data is to identify actual statistical problems specific to conservation studies and use real research questions and data as examples to explain and encourage more effective methods. Statistics is the science of analyzing real, live

data; we have tried to present it that way with a minimum of artificial, made-up examples.

The technical level of presentation varies, but this report is generally aimed at the conservation scientist who has had some training in how to conduct scientific research and had an introduction to statistics. An attempt has been made to keep most of the report comprehensible to the general conservation reader who may not have any statistical background but who is interested in the improvement of conservation research. In spite of this, a few sections will require some statistical sophistication to be fully understood.

A glossary of statistical terms at the end of this report may help the reader who has either never encountered or has forgotten some of the basic concepts needed. However, this report is not intended to serve as a textbook for teaching how to carry out each statistical method discussed. There are no references to the statistical literature. Conservation papers used for specific examples are cited.

Statistical Analysis

Our reanalyses are performed using BMDP, a statistical software package originally developed for use in biomedical research (Dixon 1985). The BMDP package was selected because we are familiar with it. Also, it is widely available, runs on most computer systems including many personal computers, and contains the full range of statistical programs required for conservation research problems. The data and BMDP setup files used for this report are given in the Appendix. The setup files can be modified to do similar analyses of other conservation research data.

Major Findings

Research Categories

Art conservation research projects and the resulting papers were easily assigned to categories of "phase" (composition, deterioration, and conservation) and "type" (method, case study, general study with real or simulated materials, and essay) as developed for this study and discussed in detail in various places throughout the report. These categorizations proved to be useful in arranging the analysis and discussion of statistical methods. Perhaps this way of thinking about conservation research could prove useful for other purposes, such as planning research or organizing the results of several studies.

Composition

Composition papers often fail to state what population was sampled and what sampling strategy was used. Clarifying these aspects of a study design should improve the conduct and interpretation of conservation research concerning the composition of art materials.

There are several specific areas in composition analysis that would benefit from statistical research on how to better use data that are

presently being collected. Among these are X-ray diffraction, lead isotope analysis, and palette composition. These are all discussed in detail in the body of the report. What they have in common is that they produce data matrices with a particular structure. The rows represent objects; the columns represent "elements." The entries in the matrix represent either the presence, amount, percent, or fraction of each element in each object. This type of data matrix also occurs in geology and ecology (species and sites) but is less common in mainstream statistical applications.

Deterioration and Conservation Experiments

Although some statistical analyses of art conservation research data have been published, they have rarely been carried out effectively. In experimental work on deterioration and conservation methods, the two critical problems are (1) determining what the experimental units are, and (2) differentiating between grouping factors and repeated measures. Because researchers are not aware of the importance of these two problems, incorrect statistical analyses result.

Studies of environmental effects on deterioration and conservation treatment effects on preservation and restoration have the structural similarity of investigating whether external agents affect art objects. The usual question is whether different agents make any difference. This is a primary application of statistical hypothesis testing. However, our survey shows that this technique is rarely used in conservation research experiments. While hunting blindly for "significant" values can be overdone, so can the opposite of ignoring hypothesis testing.

One rarely finds an article in biomedical journals presenting experimental work that does not have a test of some sort. Statistical testing allows one to separate treatments that work from those that do not. This is particularly important in conservation research where, as in medical research, we are most often dealing with probabilities rather than deterministic situations. Hypothesis testing through statistical analysis is a basis for modern medicine and agriculture. Although this study has shown that hypothesis testing is rarely used in art conservation research, judging by its usefulness to other fields we believe that it could be of great benefit to this field as well, and would allow more effective identification of optimum treatment materials and methods for the conservation of works of art.

Conservation Treatments

Medical researchers and biostatisticians have developed a progression of protocols for studies on human subjects, which are only begun after animal and laboratory experiments suggest that a new treatment is probably safe and possibly useful. The first stage is to determine whether the treatment is safe for humans. For drugs, initially small and then increasing doses are given to healthy subjects who are monitored for deleterious effects. The second stage is to work out an apparently effective procedure and dosage on small groups of actual patients. The third stage is a rigorous clinical trial of the new treatment against a placebo control or existing standard. As much as possible, the patients as well as the doctors administering the

treatment and evaluating its results are kept blind as to which patient receives which randomized treatment. This eliminates bias and makes the results much more convincing than ad hoc case histories.

In conservation treatment studies the equivalent of laboratory and animal experimentation is work with simulated art objects that are of no value other than the cost of materials. The three stages of human medical studies also have possible analogs in experiments on real art objects. Although it would often be difficult to keep the conservator unaware of what treatment he/she was applying, a defined protocol can be established, the assignment of treatments to objects can be randomized, the treatment effect evaluated by another conservator who did not see the treatment applied, and the results analyzed by proper statistical techniques.

Clinical trials are an essential component of modern scientific medicine. The statistical aspects are a subject of continuing research. There were no reports of analogous conservation trials in the work reviewed, but we recommend that they be incorporated as part of the development of modern scientific conservation practice.

Generalization

No one study can give the complete answer to any major conservation research question. The typical pilot study reporting isolated, one-time results that are not followed up do not lead to general inferences. To make generalizable statements in conservation research, such as what causes pigment fading under various conditions or what factors are involved in stone deterioration, then more sustained and long-term programs of scientific research are required. Such programs should generate multiple data sets, collected with consistent or at least compatible sampling strategies, that can be analyzed by consistent statistical methods both separately and together.

Statistical Education

Statistics are not being used very well in conservation research, but they would be useful for at least half of all published studies. Some improvements can be made immediately. For example, it is not difficult, actually saves space, and greatly improves the clarity of an experimental description, to substitute "15" for "a number of when discussing how many samples were treated. Many other easily applied suggestions are scattered throughout this report.

For some purposes, greater statistical sophistication on the part of conservation researchers is needed. A manual on the design and analysis of conservation experiments (based on a case study approach), training in the basics of using statistical software, and guidelines for conservators and conservation scientists on how to effectively get help from statisticians should all be helpful.

Statistical Consultation

As far as we could tell, only one of the papers reviewed had a professional statistician *as* an explicit collaborator and coauthor. A couple of authors acknowledged some help from a statistician and a few others gave a

statistical reference in their bibliography. There were probably other contacts either not reported or missed by us, but we have the distinct impression that there has been very little involvement of professional statisticians in art conservation research.

The active collaboration and participation of professional statisticians is needed for improved statistical analysis in this field. This collaboration should begin as early as possible in the course of a project, preferably when the experiment is being designed and before any data are collected. This collaboration is needed for three primary reasons. First, the statistical analyses currently being attempted in conservation research are not being done as well as they could be. Second, there are known statistical techniques that could be but are not now being applied to conservation problems. Third, there are areas where applied statistical research is needed, as discussed in this report, in order to develop new approaches and to fit statistical techniques current in other fields into conservation research.

This report is a joint project between a statistician and conservation researchers. It exemplifies the collaboration we strongly recommend. Perhaps an analogy will clarify the relationship we are suggesting.

In the practice of art conservation there are several possible divisions of labor between the art collector or curator and a professional, trained conservator. At one extreme, the collector can hand his collection over to a conservator and have no further involvement with the preservation and restoration of his objects. However, insufficient communication at the commencement of a restoration project may lead to unhappiness with the results. At the other extreme, a skilled amateur can attempt to perform restorations himself and never consult with a conservator. This may lead to immediate disaster or to subtle damage that may not show up for years. The latest techniques may be unknown to such a person and he may repeat mistakes for which solutions are already known. In between these extremes, the collector can learn some of the basics of conservation and be responsible for maintaining a proper environment, protecting the objects, and even performing some minor procedures, all with guidance as needed from a conservator, while leaving major procedures to the professional.

Even when a conservator is engaged, there are the extremes of beginning at the time of purchase versus waiting until the piece is essentially beyond repair. We believe most conservators would agree that earlier rather than later involvement is preferable.

The relationship between conservators and conservation scientists such as chemists is similar and can run between the same extremes of involvement and timing. Both conservators and conservation scientists may be employed as in-house staff, outside consultants, or be paid to do specific analyses or projects.

Similar again is the potential relationship between conservators and conservation researchers on the one hand, and statisticians on the other. The extreme of turning all data analysis over to statisticians is not financially feasible and may result in analyses that do not serve the purpose intended. The statistician needs communication and cooperation from the outset of a study in order to understand its purpose and goal and to contribute to its design. Knowing what was actually done rather than just what was intended is necessary for deciding on the best methods to

analyze the data. The other extreme of proceeding without any guidance from experts beyond an out-of-date introductory course in college has the same dangers as amateur restoration work of making avoidable mistakes and vitiating the efforts and outcome of the project.

We are trying to make two points here. Any argument for ignoring statisticians can be turned into an argument for conservators to ignore chemists and collectors to ignore conservators, with about equal validity. On the other hand, when chemists and conservators do decide to consult with a statistician, they might consider their experiences on the other side of the fence for some guidance on how to proceed to make the experience as fruitful as possible.

Chapter 2

Composition of Art Materials and Objects

Organization

Studies pertaining to composition were subdivided into the specific types listed in Figure 1. This table gives the number of papers of each type in each journal. Case studies are limited to one or a small set of objects, without regional or chronological generalizations. General studies focus on a regional or chronological group of art objects. Determination procedure studies develop and present methods for identifying the materials of which art objects are made.

Figure 1.
Frequency of art composition studies

	Journal				
Study Type	SC	JC	TB	PP	ALL
Determination Procedures	10	4	1	3	18
Case Studies	12	2	16	5	35
General Studies	6	2	1	8	17
All Composition Studies	28	8	18	16	70

Our discussion of the statistical problems exemplified by each of these three types of composition studies is contained in a separate section of this chapter. In general, the presentation for each type or subtype begins with a discussion of the goals specific to studies within that type and the statistical considerations and procedures particularly pertinent to such work. This is followed by one or more examples taken from the literature reviewed. The examination of each example typically includes a succinct description of the study and data collected, a presentation of how we would approach the analysis, and finally a summary and critique of the author's methods. A general discussion of other papers in the group is sometimes included.

Composition: Determination Procedures

Validation

The typical goal of papers in this category is to present a method for determining the composition of art objects that readers can apply in their own work. A major statistical problem associated with these studies is to validate that the method works. A complete verification has three steps: Get correct answers (1) on the training set, (2) on new samples, and (3) by users other than the developers.

Step 1

The first validation step is to demonstrate that the method can give correct answers when applied by the investigators to the original sample material.

For quantitative measurements, such as the atom or weight percent of elements in a stone or metal sample, this can be done by presenting a scatter plot of the value resulting from the new method versus the value resulting from an accepted standard method or a known true value. Alternatively, the same data can be presented in a tabular format. Either way, a correlation coefficient is calculated and shown to be sufficiently high for the purpose of the measurement.

Many analytical methods in conservation, however, are concerned with qualitative determinations. An example is the identification of the pigments in a painting. We can do essentially the same thing with such categorical data as with quantitative data. Instead of scatter plots and product-moment correlations applicable to numbers, we can substitute two-way contingency or cross-tabulation tables and correlation measures designed for categories.

Step 2

When the method being presented is explicitly designed to give the correct answer on all the training samples, the first step is not applicable. There still remains the problem of showing that it will work on new material and when applied by new people. There is precedent for this in other fields. For instance, a biologist, after writing a plant or animal identification key that works for the specimens considered to be prototypical examples, may make both tests, with both new specimens and other biologists.

Cordy and Yeh (1984) present a flow chart for the identification of three blue dyes (indigo, Prussian blue, logwood) used on nineteenth-century cellulosic fibers. The procedure outlined in the flow chart was developed as a result of a literature review and of original laboratory work in which flax thread samples were prepared and dyed using nineteenth-century recipes and processes. Some samples were artificially aged, and the dyes were analyzed in both aged and nonaged samples. An acid digestion technique was used to release dyes from fibers into solution, then UV-VIS spectra, IR spectra, and wet chemical analyses were recorded and examined for discriminating features to be incorporated into the flow chart.

This flow chart presumably gives the correct answers on the training set. It could be given the second and third step of validation by giving new samples with known dyes to a new analyst who would attempt to identify each dye correctly by following the procedures outlined in the flow chart. In this type of test it would be important to code the samples in such a way that the analyst did not know the identity of the dye. The test samples should include real samples from historical objects that have been analyzed by the older, more laborious method. Real samples often cause more difficulties and problems than synthetic laboratory samples.

A problem that could use more research is how to decide when a sample does not fit into any of the categories allowed by the identification procedure. There may have been at least a fourth blue dye used in the nineteenth century.

Indictor, Koestler, and Sheryll (1985) studied the detection of mordants through scanning electron microscopy with energy-dispersive X-ray spectrometry. SEM-EDS is an established method already validated on other types of samples, so this is effectively a Step 2 validation study.

Twelve cochineal-dyed wool samples were mordanted with known preparations, then submitted without identification for analysis to test whether the technique could qualitatively determine the metallic elements of the mordants. All twelve analyses gave a clear identification of the mordants used, although this was somewhat difficult for the reader to see since the analysis results were in two tables and the mordant composition in a third.

Steps 1 and 2

Among the procedural papers dealing with art material composition, Step 1 and Step 2 validations can be found. For example, Allison and Pond (1983) used known technical information about bronze casting and duplication methods to derive a procedure for identifying bronze statue copies, using internal measurements and shrinkage data. Their method for tracing several generations of copies back to the original wax model was refined during the course of their example problem, which was to identify duplicates of a model by an Italian Renaissance sculptor as being either from the same (possibly original) model, or as being casts from a bronze model. Although it would probably have been better to use objects with a well-known history, the authors felt that this was a basically straightforward and unquestionable example. Thus, their example problem can be considered a Step 1 validation. The fully refined method should have been further validated by applying it to another example.

An example of both the first and second stages of validation is found in a paper by Jan Wouters (1985). He developed a method to quantitatively determine red anthraquinone dyes on textile fibers using high-pressure liquid chromatography. He first demonstrated that the method works, using pure dyes extracted from plant roots and insects. He next demonstrated that the method can work on actual textile samples, using modern textiles that he dyed himself with the same known materials already analyzed. Finally, he analyzed ancient textiles with previously unidentified dyes.

Step 3

There are no papers in the conservation literature surveyed that explicitly carried out a Step 3 validation. Nor were there any attempts to validate previously published conservation research techniques.

One example of a technique for which a proper validation study could be particularly useful is pigment identification by optical microscopy. An evaluation of the degree of reproducibility of identifications between different analysts is especially important for such a widely used conservation research technique that depends upon qualitative assessments.

Composition: Case Studies

Composition studies published in conservation journals are carried out for three reasons and can be grouped accordingly:

1. To determine composition as an end in itself (the corresponding papers are usually case studies)
2. To answer art historical questions as to authenticity and provenance (this typically applies to general studies)
3. To decide on the most appropriate conservation treatment (studies measuring composition for conservation reasons are included in Chapter 4, "Conservation Treatments and Materials")

Sampling within an Object

In this subsection we discuss the statistical problems related to determining the composition of a single object. Problems of sampling between objects, rather than within a single object, are deferred to the "Sampling Groups of Objects" subsection of "Composition: General Studies" (this chapter).

The goal of sampling within one object is to determine the list of components and sometimes an average quantitative measure for each. Traditional discussions of sampling cover the twin questions of how samples should be selected (the sampling strategy) and how many should be chosen. In composition sampling there is the additional question of how large each individual sample should be.

There are at least six possible sampling strategies:

1. Analyze the *entire object* instead of choosing just a portion. Examples are X-ray radiographs of paintings and statues. When possible, this is often the best method but, for destructive analyses, usually impossible.

2. *Homogenate* (grind, powder, dissolve, etc.) the entire object, and sample and analyze a portion or aliquot of the result. For example, hunks of copper slag are often powdered and a standard amount of the powder analyzed by X-ray fluorescence. Again, however, this strategy is essentially impossible for art objects.

3. Take *randomly located* samples within the intact object. This means selecting points determined by numbers from a random process (throwing dice, flipping coins, drawing well-mixed slips of paper), random number table, or computer pseudo-random number generator and does not refer to the typical arbitrary or haphazard sampling often mislabeled by the term "random." This strategy is effectively equivalent to strategy 2, which brings multiple random points together into the portion actually analyzed. It is usually more complex to carry out than strategy 2 but conservationally more acceptable than destroying the entire object. In either case the true values for the object are estimated from a portion, and the estimates have known statistical properties. As long as the area available to be sampled is larger than the sample to be taken, this strategy is applicable.

4. Choose *regularly patterned* samples. This usually means taking samples at equal intervals across the object. This strategy is sometimes easier to execute than strategy 3, but the danger is that if the object has spatial structure at the same scale as the sampling interval the result may be very biased. Systematic samples of a city at block-sized intervals could give the impression that the city is all asphalt, concrete, metal, glass, grass, or wood, depending upon where we start with the first point.

However, if one is looking for spatial patterns, then systematic sampling is advantageous if the sampling interval is small enough.

5. *Haphazardly* or *arbitrarily* select points. This includes restricting samples to particular positions for aesthetic or other reason extraneous to the immediate goal of composition determination. This common strategy has the danger of giving a biased result. It may be the easiest procedure, but gives no basis for generalizing from the sample to the entire object. If it is the only strategy possible, then it is better than the strategy of no sample at all.

6. *Intentionally* select or sample components not yet examined. This is a typical strategy for palette studies, and may be necessary if one is trying to identify all the rare components of an object, which might be missed by a random sample.

Choosing a Strategy

From a statistical viewpoint, if one wants to know the composition of a particular object, complete analysis is best, and both random and regular strategies are superior to arbitrary sampling for obtaining a statistically accurate estimate of the average composition of the object.

If complete analysis is not possible, one should take multiple samples within the object. While the first answer to "How many?" might be "The more the better," there is a point of diminishing returns that sets an upper limit to the number needed. The number of samples to take would depend upon how accurate one wants to make the estimate (what size of confidence interval is acceptable). The number necessary will also depend upon the type of objects being sampled and their degree of heterogeneity.

Constraints within the field of conservation research often preclude large numbers of samples. One is often very lucky to be able to take one sample from an object. It would be desirable if this single sample were selected at random, or failing that, by some consistent criteria relevant to the measurement of composition. Even this is difficult when the sample must be from a hidden location not visible to viewers of the piece.

If only one sample is taken from an object, then one can only make direct conclusions about the composition of the particular point sampled. To extend this to the object as a whole requires some assumptions. If there is no systematic relationship between composition and convenience of sample location, then the composition of the sample is an unbiased estimate of the composition of the object as a whole. Thus, to obtain an average composition with one sample, one must be certain that any variations are not systematic or make an assumption about what those variations are.

If two or more samples are analyzed from the same object, the results are more likely to be representative. Even with two, it becomes possible to make an estimate of the variability between samples, and therefore of the accuracy of the composition estimates. We therefore recommend, especially for case studies, the analysis of at least two *independent* samples. If we have multiple samples from similar objects, and we assume that the variability in the current object is about the same as in others of its type, then we can also say something about how good our estimate is.

When it is not possible to take even two samples, the size of the sample analyzed becomes very important. In physical objects, sample size is a continuously variable entity. Composition variations at scales smaller than the sample will tend to be averaged out while large-scale variations will lead to bias. Therefore, large samples will average across a large range of variation while microanalytical techniques will be vulnerable to microvariations.

The decisions about sampling from continuous but heterogeneous entities include determination of the method, number, and size of samples to be taken and analyzed. For each aspect, there are cost and benefit trade-offs. For a major project or series of studies of a similar type, a model can be constructed that will make some of these explicit and allow a more rational choice.

Examples

In some of the case studies surveyed for this report it is clear that the sampling procedure was designed to intentionally select specific components (strategy 6 above). Generally, however, the sampling strategy is not discussed and so we cannot assume more than that sampling was carried out haphazardly at arbitrarily selected points (strategy 5 above).

The primary reason for authors to identify their sampling strategy is that it helps the reader to evaluate the results presented. For example, Marchese and Garzillo (1984) studied the chemical and physical characteristics of the tesserae materials in the wall and floor mosaics of the Cathedral of Salerno. Fourteen tesserae from cathedral mosaics were analyzed, along with one sample from a mosaic in Pompeii for comparison. Three samples were taken from the cathedral floor and the remainder from three different mosaics now in the cathedral museum. Analysis included a visual color determination using Munsell color standards (for hue and value/chroma), specific gravity and hardness tests, mineral analysis by X-ray diffraction, and qualitative elemental analysis by scanning electron microscope with energy-dispersive X-ray fluorescence. No mention was made anywhere in the paper about how the 14 samples were selected for analysis. Thus we cannot judge whether these samples represent the full range of mosaic materials existing on the cathedral, or whether they are only the most commonly occurring materials, or ones that stand out in some way that would make them most likely to be selected.

Sack, Tahk, and Peters (1981) researched materials and painting techniques used to create a painting ascribed to third-fourth century A.D. Egypt. A macroscopic examination identified the overall structure of the painting; microscopic and microchemical tests were done to identify the canvas fibers and the pigments, with ammo acid analysis to identify the adhesive used to attach the canvas to the fabric beneath and the binding medium used for the pigments. The authors illustrate where the sample sites are located, but never mention how and why those sites were selected.

Rodriguez, Maqueda, and Justo (1985) asked: What materials and firing temperatures were used to construct the terracotta sculptures from the Seville Cathedral porticos? They applied six methods of technical analysis to an unknown number of samples. It is not clear whether the different analytical methods were applied to the same or different samples.

An intentional sampling strategy was followed by Stodulski, Farrell, and Newman (1984) in their study of the range of pigments used at the Persian sites of Persepolis and Pasargadae. They apparently sampled a small amount of any appropriate (relatively uncontaminated) painting fragment encountered on the limestone reliefs at the sites. All samples were analyzed by X-ray diffraction, qualitative X-ray fluorescence, and Fourier transform infrared spectrophotometric techniques. In addition, they mention that optimal and minimal sample sizes were determined for these specific materials and analytical methods.

Palette Studies

The most common type of composition case study is the palette study of one or a few paintings of a particular artist, school, or culture. When nondestructive qualitative estimates of composition are made, such as in pigment studies with energy-dispersive X-ray fluorescence, one has the option of random, regular, haphazard, or broad-spectrum selection strategies and even combinations thereof. If material is removed from the painting, the sampling will be more constrained. It should be clear whether the goal is to select the more common pigments, those of a certain color range, or all pigments used in any quantity. Making the sampling goals and procedure clear will help the reader to properly interpret the results given.

In a technical study of Hogarth's *Marriage à la Mode,* Ashok Roy (1982) gave some details about his sampling method. Samples were taken from all six paintings comprising this work. Irregular painted edges concealed by the frames allowed relatively many samples to be taken along the edges, while samples removed from the main picture area could only be removed from sites of old flake losses or at the broader surface cracks. A total of 70 samples were removed from the 6 paintings for X-ray diffraction and laser microspectral analysis. The pigments found in each painting were listed separately, and a summary of the total palette discussed in light of painting information found in various historical texts. Because the goal of the project was to compare the total palette composition of the six paintings with published accounts of contemporary painting methods, we might assume that the selection of samples was intended to represent all hues and pigment types, but this was not stated.

If an estimate of the relative abundance of the different pigments was desired, random samples could be selected from the range of acceptable sampling sites (edges under the frame, existing flake losses, and surface cracks). The palette estimates would then have known statistical properties for comparison. In general, whatever area is both available and relevant to the particular study can be randomly sampled.

An explicit broad-spectrum sampling procedure was used by Calamiotou, Siganidou, and Filippakis (1983) to find what pigments were used on a wall painting of a house of the first Pompeiian style (400-168 B.C.) found in Pella, Greece. They analyzed 24 samples of 8 different colors and included samples of 3 plaster layers. Analytical techniques used were X-ray diffraction and a qualitative elemental analysis by X-ray fluorescence. They explicitly stated that they had sampled to represent all pigment hues: red, green, light-blue, white, yellow, grey, black, and pink. Except for the pink, they collected at least two samples of every hue, and so increased the chance that the full range of pigments used for each hue would be represented.

X-ray Diffraction Data

X-ray diffraction data come in the form of a diffraction film, spectrum, or set of Angstrom spacings or d-values (the last two can be derived from the first two). About 3% of the space in *Studies in Conservation* is occupied with raw X-ray diffraction data, along with the JCPDS reference patterns used to identify specimens. In comparison, approximately 0.5% of the space in that journal is devoted to reporting the results of statistical analyses. While attesting to the importance of this analytical technique to conservation research it is unusual to devote so much space to raw, unreduced, analytical data. It certainly seems unbalanced to devote six times as much space to this one type of raw data as to all statistical analyses. This is perhaps the only case in which we consider that too much rather than too little of the data is being published.

If the diffraction pattern matching procedure is objectively standardized, there should be no need to present the raw instrumental data, any more than with other techniques. If, on the other hand, diffraction pattern matching is so subjective and idiosyncratic that researchers feel compelled to publish d-values and measured intensities next to those of the reference patterns so others can evaluate the match and decide for themselves whether or not it is a correct identification, then there is a need to develop standardized, generally known and accepted, matching algorithms.

Although there are numerous complications that can arise with diffraction pattern matching, including problems with orientation effects, differences in equipment used, and variations in the skill of analysts, it is still possible to give a quantitative numerical assessment of the closeness of fit of a sample spectrum to a reference spectrum. The complications mentioned above can be taken into account when interpreting matching coefficients. If local variations are a major problem, then comparisons should be made against local rather than published reference standards.

For example, Orna and Mathews (1981) give d-values for samples and reference standards of the commonly used mineral pigments lazurite, lead white, vermilion, orpiment, massicot, and lead-tin yellow in their Tables 2, 3, 4, 7, and 8. Although the tables are titled as a "comparison" of the appropriate d-values, no comparison measure is given. The d-values for a sample are simply listed next to the d-values of a reference specimen, and it is left up to readers to do their own comparing.

For the more common, easily identified minerals, it may be enough to simply state that they were identified by the X-ray diffraction analysis. If one wants to list the d-values of a sample and reference, a quantitative measure should be used to compare the d-values of the two rather than leaving it to the reader to visually assess or mentally calculate.

Comparison Measures

In the strict sense, similarity measures are the converse of dissimilarity or distance or difference measures in that one goes up while the other goes down. Since either type can be changed to the other by changing the sign, we will use similarity measure as a generic term for either type.

A possible similarity (distance) measure for a diffraction pattern or spectrum is the integrated squared difference. This measure will depend upon differences in both peak intensity and location. A value of 0 repre-

sents perfect similarity or identity. Analogous measures can be generated by d-values. A simpler measure is the mean fractional error for each peak that is present in both lists. It should be recognized that in matching unknown with known lines, agreement of relative intensities of corresponding lines might also be significant. Measures can also be developed for comparing peak intensities.

As an example, Figure 2 reproduces d-values for a lazurite sample analyzed by Orna and Mathews and the standard reference values. In order to compare the two specimens quantitatively, the raw difference for each peak is divided by the reference value to obtain a relative error.

Figure 2. Quantitative comparison of lazurite d-values (data from Orna and Mathews 1981:65)

d-value			
Reference	Sample	Difference	Relative Error
6.43	6.18	.25	.039
4.54	4.50	.04	.009
3.71	3.77-3.65	0	0
2.87	2.99-2.86	.06	.021
2.62	2.64-2.60	0	0
2.27	2.27	0	0
2.14	2.12	.02	.009
1.78	1.77	.01	.006
1.66	1.67	.01	.006
1.61	.60	.01	.006
1.56	.55	.01	.006
1.51	.50	.01	.007
1.47	.47	0	0
1.37	.36	.01	.007
1.31	.31	0	0
1.28	1.28	0	0
1.24	1.24	0	0
		Mean value	.007

The raw differences need to be inversely weighted according to the expected magnitude. This can be done by dividing the difference by an error estimate (such as "sigma," the standard deviation of repeated measurements) to get a normalized number ("Z" if sigma is used). Any number proportional to the error will have the same effect as to weight. In Figure 2 the reference value is used as a crude estimate of the relative magnitude of the expected error since this is true for many instruments (hence the widespread use of relative versus absolute error) and close enough here for illustrative purposes. This relative error, although still an approximation, is an improvement over raw differences in the present example. Lines that differ by a factor of 25 (.25 to .01) in raw difference scores for d-values differ in relative error by only a factor of 6-7 (.039 to a mean of .0063), and others differing by a factor of 4 (.004 to .001) are nearly equalized (.009 to .0063). More refined error estimates would require analysis of empirical results for many samples or a theoretical analysis based on the principles of diffraction.

If there is some reason to question the correctness of an identification, one could make use of the quantitative similarity measure discussed above by showing that the mineral identified as matching is quantitatively closer than other potential matches. This would give a standardized criteria for matching that could be summarized briefly. A possible result might be: "All 15 lines of reference A match the observed sample lines

within 1%; for the best potential alternative, mineral B, only three lines match to the same degree."

If quantitative comparison measures were used, particularly if peak intensities were included, apparent mismatches would stand out more than they do when d-values are simply listed for mental comparison by the reader. A poor match could be due, among other things, to either a deficiency in the pattern or the presence of another unidentified mineral.

In the conservation literature reviewed no use is currently made of quantitative comparison measures for d-values, intensities, or X-ray diffraction spectra as a whole. We believe that this is a subject that would be worth further research.

Composition: General Studies

Sampling Groups of Objects: Authentication and Provenance

Inference

Where authentication or provenance is the goal of a composition study, statistical inference is always used, even if only implicitly. The important questions involved are how many objects are necessary (in addition to how many samples within an object), and how does one make inferences and put confidence limits on the results?

The rationale for all sampling strategies is that the inference mechanism and all probability statements used in making an inference are based upon a mathematical model of how the data are gathered. The validity of these probability statements in reference to real data depends upon the validity of that mathematical model in relation to the real sampling process.

The following five steps are the basis of statistical inference:

1. Gather data.
2. Construct a mathematical model of the data gathering process.
3. Derive probability statements from the model.
4. Assume that these probability statements at least somewhat correspond to probabilities in the actual data gathering process.
5. Infer the nature of the unsampled universe from these probability statements.

Random Sampling

Random sampling, whether from the objects of interest taken as a whole, or from predefined strata, is a commonly used sampling model. It has the advantage of making the probability calculations easy to carry out. Unfortunately, it is often very difficult or impossible to use this model when working with art objects.

For example, if one has 200 statues from a particular region available at a museum, but sample removal and expensive analysis can realistically be carried out on only 20 of those statues, a possible random sampling method would involve putting 200 slips of paper with identification numbers into a hat and blindly drawing 20 slips after thorough mixing. One can then make reliable inferences about the total group of 200.

Suppose, instead, that there are 100 objects at each of two museums. A random selection of 20 could be drawn from the 200 objects. But if the objects from the two museums were expected to be different or the two curators each put a limit of 10 samples from each museum, then we would select 10 from the 100 at each museum in a separate selection process. This would be a stratified random sample.

A compromise method used in other fields such as biomedicine, which also has practical constraints on sampling, is to take what you can get. However, then the researcher should restrict inferences made in Step 5 to the population actually sampled rather than the population he or she would have liked to have sampled. It is important then to describe the objects actually available for sampling and the method, if any, for selecting the subset.

Koestler, Indictor, and Sheryll (1985)

They analyzed 13 fibers from 7 different silk textiles for metallic mordant elements by SEM-EDS using modern textiles with known mordants as standards. The textiles, all from a group known as the *Buyid Silks* said to have been excavated in Persia in 1925, are attributed to the ninth-tenth century A.D. The authors admit that they cannot authenticate the textiles with the data obtained, but claim that the mordanting materials are "consistent with those found on ancient textiles."

The experimental design and its description could be improved in several ways. First, it should be clarified how the seven textiles were selected for analysis—is this the complete set of *Buyid Silks* available, or is this a selection (haphazard?) from a larger collection? Second, the only information given about the comparative ancient material, that it is "Eastern Mediterranean," is from the title of the relatively obscure conference report. The substantive results of this comparative material should at least be summarized. Third, we need some evidence that this comparison group has some relevance to authentication of tenth-century Persian silks. Fourth, without determining the full range of modern as well as ancient mordanting procedures, we cannot rule out that these data are equally consistent with modern materials. Fifth, the result of their C-14 analysis should be given rather than dismissed as "uninterpretable."

Generalization

When one studies a haphazard collection of objects it is difficult to know how far to generalize the results. In medical trials it is usually considered desirable to keep a log of all patients who meet the basic criteria of having the disease under study but are excluded from the trial for various other reasons. This allows statements to be made about the excluded patients: their frequency, reason for exclusion, and similarity to those selected. These factors are important evidence as to how far beyond the group studied the conclusions of the trial apply.

An additional factor important in determining the extent to which inferences can be made is the fact that the observed variance between sampled units must reflect both the true between-unit variance and the within-unit variance (variance of repeated samples within each unit). For

example, in a t-test the crucial denominator, which should be as small as possible, is the ratio of the observed standard deviation to the square root of the sample size. This can be decreased either by increasing the number of objects sampled or by decreasing the standard deviation or variance of the measurements for each object. One can reduce the expected observed variance toward its lower limit of the true between-object variance by making the individual measurements more accurate. For any particular study it would be useful to develop a model of relative cost versus relative benefits of sampling more within each object or of increasing the number of objects included.

As a general rule we can say that if the number of objects sampled is relatively small (such as 10) it will probably be more valuable to sample more objects rather than more intensively within each object.

The final conclusion we can make from this discussion of random, regular, and haphazard sampling is that doing real science on idiosyncratic heterogeneous objects is difficult at best, and that good statistical work under these conditions is very hard. However, major improvements can be made by noting what population one is actually sampling from, why the particular specimens analyzed were selected, and what the justification is for the sample size. It will then be much more clear to what extent inferences can be made beyond the specimens actually analyzed.

Spread Sampling

Spread sampling explicitly attempts to encompass as much of the actual variation as possible. In an authenticity study, the logic may be to. exclude the possibility of a piece either being old or being modern by showing that it has a characteristic never found in one of the two groups of objects and sometimes found in the other, so sampling to get all the possibilities in each group may be the most useful. The associated probability statements can take the form of giving the chances of having missed something actually present in either group.

Palette Studies

Sampling for variation applies both within and between objects. In both cases, palette studies are the most common application of this strategy. And in both cases, a primary question is, "When should we stop; when have we looked enough?"

Investigators in ecology have studied the relationship between the cumulative effort that has gone into looking for new species within an area and the number found. Palette studies that appear in the conservation literature for a particular artist, region, or time period could benefit from such a cumulative effort analysis. How well one can determine whether or not a pigment is consistent with the palette under study depends upon how much work has gone into finding the possible choices. This will be particularly true for minor pigments and accessory compounds. Ecological studies show that one can project the total number of species present from the various numbers found at various levels of effort. Thus for any particular palette study one can keep track of the overall effort that has been made and continue to collect results until the effort-result curve levels off enough to make it no longer cost efficient to continue collecting analyses. For any particular project one can stop collecting new data at whatever

point of probability one considers desirable (the probability that you may have missed a particular number of pigments that should be included in the palette).

Palette studies could also make use of a stratified application of the principle of diminishing returns. Most palette studies appearing in the conservation literature include an analysis of only one sample of each color found on a particular painting. The implied assumptions are that artists use the same pigment for each color throughout an entire painting, that colors now similar after fading and deterioration were similar upon application, and that artists and samplers all discriminate colors the same way. If two samples are taken of each color and each pair are found to consist of the same pigment, then such an assumption would be demonstrably reasonable. If enough paintings by the same artist have been analyzed to show that this principle appears to hold true for that artist, then it would be reasonable to begin to analyze only one sample of each color per new painting studied. But unless there is such a data-based rationale for assuming a one color-one pigment relationship, palette studies could be improved in the area of statistical inference by analyzing two samples of each visually distinguishable color. This improvement in the research design would allow one to more reliably examine changes in a palette over time or between artists, as it would allow one to compare with more certainty the consistency within a specific painting versus across paintings.

Orna and Mathews (1981)

They mineralogically analyzed pigments from the Glajor Gospel book at UCLA to compare the materials used by artists of two separate but nearly contemporaneous workshops and to compare those workshops to others in Byzantium and western Europe.

Five different painters of book illustrations from two different workshops were identified within the Glajor Gospel book on the basis of style and working methods. Seventy-six samples representing the hues used by each artist were mineralogically identified by polarized light microscopy and X-ray diffraction. The hues used by each of the five artists and the mineral pigments used to achieve those hues are listed, as well as the total palette of the book. However, they found almost no published data for comparison.

A positive feature of this study, relative to analytical studies that merely list composition data, is the examination of art historical questions with the pigment compositions. This endeavor could have been improved further by the application of clearer hypothesis testing methods.

The key point is that the groups were defined before any samples were taken. The starting hypothesis is that each of the five artists and the two workshops can be distinguished on the basis of working method, including pigment choice. Data are then collected to confirm this.

With such a hypothesis, it is necessary to determine prior to interpreting the data (and preferably prior to its collection) what the rules of corroboration will be. What criteria will support or refute the hypothesis? In this case, what defines significant differences between palettes? A *post hoc* selection of favorable evidence and ignoring of other evidence is not very convincing.

The authors claim that their evidence supports their hypothesis. Another reader of the data table could read the results differently, and arrive at another conclusion. One alternative reading of their pigment results indicates that equal support can also be found for the existence of four artists in one workshop and a separate solo artist. This alternative hypothesis could be supported by the fact that four artists use gold hues and one never does; that same artist also achieves a magenta hue by a different method than the other four do.

An alternative method for undertaking a project of this type would be to first define the experimental unit—which here could be the individual paintings within the book. The hypothesis is that five particular artists from two specific workshops painted each one. Because we have a hierarchically structured hypothesis, it would be better to first split the paintings into the two workshop categories and test that hypothesis; then the problem of the existence of five painters could be separately addressed.

Similarly, a hierarchical method could be used to compare the pigment analyses of possible artist and workshop groups. First, questions could be addressed concerning the range and number of hues found for each group or artist. Secondly, comparisons could be made of whether or not they used the same pigments to achieve their hues.

In order to reliably test the identification of five painters and two workshops, it would have been better to sample the complete palette of each painter with replication. Without some replication we can never be certain about the results. For example, if a distinguishing criterion is that four artists use ultramarine and one uses azurite, and we only have one blue sample from each artist, we cannot rule out the possibility that all artists may have in fact used both pigments and chance alone caused us to sample these particular pigment choices. If two samples were taken of blue hues for each artist, and we still had the four-ultramarine one-azurite pattern, our certainty would be greatly increased.

In this line, it would have been helpful to have more information about the sampling method. How many different paintings were sampled for each artist? If all samples for a particular artist came from only one painting the inferences we can make about the artist's palette are much narrower than if a wide range of different paintings was involved.

England and van Zelst (1982)

They identified pigments from 15 seventeenth-century New England portrait paintings, most by anonymous artists. The study was intended to test the conclusions of stylistic studies which suggest that there were only a limited number of artists active in New England (Boston) in the latter part of the seventeenth century. Pigment types were determined through elemental analysis by energy dispersive X-ray fluorescence and by microscopic characteristics. They also studied the overall structure of the paintings with X-ray radiography.

The authors conclude that there is a close correlation between the pigments used in these portraits and those in use contemporaneously by European artists; and that this implies the majority of raw materials were probably imported from Europe. However, they do not list the pigments they consider to be European, reasons why those pigments could not have

been locally produced, or references to late-seventeenth-century European palette analyses. Thus these conclusions cannot be evaluated by the reader.

From the analyses of the 15 paintings (their table, pp. 92,94), they conclude that through time there is an increasing sophistication in techniques with use of a more layered structure by 1670 and an increasing range of colors used thereafter.

Their method of selecting samples was not discussed at all. One painting had no pigments listed, only a red ground. Two other paintings each had only two pigments listed. These results imply that they selectively sampled only certain hues or pigments, not including the full range of pigment choices.

There are several possible ways to analyze palette data such as presented in this study. As is usually the case when the data are in the form of a true matrix, with all values measured in the same unit and the choice of row and column arbitrary, both rows or columns could be analyzed equally well. In palette studies, the relationship of paintings to each other and the relationship of pigments to each other can both be analyzed. Statistical techniques can be used to compute similarities, correlations, or distance measures between each pair of paintings or each pair of pigments. These relationship matrices can then be analyzed either by clustering methods that arrange the entities in groups, or ordination methods which locate them in multidimensional continua instead. If we think that there are changes occurring over time, we could in addition do a regression analysis.

There are many similarity measures available, and for each particular research project some thought would have to go into deciding which would be the most useful for the specific problems under study. One can also try several and see which results remain consistent in spite of the differing details of the analysis.

To give an idea of where such analyses lead, we calculated the product-moment correlation of the presence-absence measures for both pigments and paintings in England and van Zelst's paper (Appendix A.1, Figure 3). In both cases, the items have been rearranged so as to bring the most similar paintings and pigments together. The highest correlations are located along the diagonal. For convenience, the correlations have been multiplied by 100 so they can be interpreted as percents ranging from -100 to 100, instead of fractions ranging from -1 to 1.

A positive correlation of 100% between two pigments would mean that they have the same pattern of occurrence in paintings—either both or neither would be present in any particular painting. A negative correlation of -100% would mean that they have contrary patterns of occurrence—exactly one of the two would be present in each painting. An indifferent correlation of 0 would mean that the occurrence patterns have no particular relationship to each other.

The interpretation of painting correlation is essentially the same, after the roles of painting and pigment are reversed. Identical palettes are represented by +100; contrary palettes, where each pigment is in one or the other but not both, by -100.

The reordering of the items being compared is the first step in any research as to which paintings are most similar to each other. To show that this set of paintings is as similar to European paintings as to each other,

and that the pigment range of paintings in this study are correlated with European palettes, a similar ordered matrix could be constructed that includes analytical data from seventeenth-century European paintings.

Our ordered matrix shows that the most highly correlated pigments (based on painting occurrence) are copper resinate with vermilion, green earth with vermilion, and lead-tin yellow with red lake. The most highly correlated paintings (based on pigment variety) are the portraits of Elizabeth Wensley (1670-1680), John Wensley (1670-1680), and Major Thomas Savage (1679). There are a number of other paintings with high positive correlations, but no overall chronological relationship is apparent. For example, the early portrait of Elizabeth Eggington (1664) is more highly correlated with a painting of Captain John Bonner attributed to 1690 (.45) than to the other painting from 1664 of Dr. John Clarke (-.43).

Figure 3.
Pigment and painting correlation matrices (%) (data from England and van Zelst 1982:92,94)

PIGMENT	1	2	3	4	5	6	7	8	9	10	11
1 yellow lake	100										
2 red lake	25	100									
3 light yellow	-29	46	100								
4 vermilion	16	34	34	100							
5 copper resin	-25	-20	34	56	100						
6 green earth	-22	-5	-5	49	33	100					
7 ultramarine	-7	25	-29	16	-25	33	100				
8 realgar	-7	-29	-29	16	29	33	-7	100			
9 smalt	-22	-33	-5	-12	33	17	-22	33	100		
10 umber	-13	13	13	-8	-13	-7	-13	-13	-7	100	
11 gold	-13	-20	-20	-45	-13	41	-13	-13	-7	17	100

PAINTING	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Captain John Bonner	100														
2 Captain Thomas Smith	39	100													
3 Robert Gibbs	39	54	100												
4 The Mason Children	7	26	67	100											
5 Mrs. Patteshall & Child	-4	8	54	67	100										
6 Elizabeth Eggington	45	26	67	63	67	100									
7 Mrs. Freake & Baby Mary	45	26	67	27	26	63	100								
8 Adam Winthrop V	-7	15	56	47	56	47	47	100							
9 Elisha Hutchinson/ Sir George Downing	4	8	54	26	8	26	26	56							
10 Major Thomas Savage	7	-15	26	27	67	63	27	47	26	100					
11 John Wensley	21	4	4	7	39	45	7	31	4	83	100				
12 Elizabeth Wensley	6	-15	-15	-10	26	27	-10	10	-15	63	83	100			
13 Dr. John Clark	-35	-29	-29	4	-29	43	43	4	24	4	13	4	100		
14 John Davenport	-35	-29	-29	4	-29	43	4	4	-29	43	-36	43	39	100	
15 Edward Rawson	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

To test whether there is a greater range of pigments used after 1670, we did a regression of pigment variety against time to see if there is a significantly positive slope. We did not find any increase in pigment variety over time.

Other palette studies in this survey for which similarity measures could be useful include the paper by Newman and McKim-Smith (1982) concerning the materials and techniques used by the seventeenth-century Spanish painter Diego Velazquez; a technical analysis of paintings by Jan van Goyen and Salomon van Ruysdael (Gifford 1983); a study of the materials used by Paul Cezanne (Butler 1984); and a study of the evolution of the palette of Seurat based on an analysis of his *La Grande Jatte* and smaller studies (Fiedler 1984).

Lead Isotope Analysis

There are four stable isotopes of lead (Pb) with atomic numbers 204, 206, 207, and 208. When lead is analyzed by mass spectrometry the result is four counts proportional to the number of atoms of each isotope in the sample analyzed.

The first step in analyzing lead isotope data is to decide whether and how to transform or rearrange these four pieces of information. One reason for data transformation is to isolate or bring to the forefront specific factors or aspects of the data. In this case, we would be interested in particular features of the lead isotope composition of each object.

Each isotope count is proportional to the amount of lead analyzed. In studies of art materials, the amount of lead actually measured is arbitrary and irrelevant to the purpose of the isotope separation (as long as enough is measured for good accuracy). It is therefore desirable to transform the four pieces of information so that the total amount of lead is isolated as one piece of information that can then be ignored. This is done by summing the four isotope counts. The other three numbers that are then used for analysis should be made independent of the lead total.

One way to do this is to divide each count by the sum of the four counts. The four fractions must add to 1, making them interdependent, so that any one can be derived from the other three, leaving exactly three independent pieces of information.

Another approach is to take ratios of the raw counts or, equivalently, of the derived fractions. Any three ratios, out of the twelve possible, that are not reciprocals of each other, can be used as the three pieces of information that are independent of the total count. Typically, the three ratios used are the ratios of three of the isotopes to the fourth. Both Pb 204 and Pb 206 have been used for the denominator. Some instruments are set up to output these ratios directly.

There is an advantage to using ratios under some circumstances. An example is when one is comparing results from different studies for which different sets of elements have been analyzed, or in cases where the total fraction is not known. In these cases measurements for some elements can be used in the form of ratios even when fractions of the total are unavailable.

However, this problem is not applicable here, as we always have the full fractional composition of lead. Typically, the distribution of ratios is apt to be less desirable than the distribution of fractions for the purpose of statistical analysis. Therefore, in the absence of compelling reasons to

rely on ratios, we are more likely to be successful in any given statistical analysis if we use the fractional data.

Example

Among the articles reviewed, the only example of lead isotope data is the appendix by Brill, Barnes, and Murphy to the article by Lefferts, Majewski, Sayre, and Meyers (1981:32-39). The article reports technical examinations of the classical bronze horse in the Metropolitan Museum of Art made for the purpose of authentication. The appendix presents lead isotope data for two samples from the original casting of the horse, two samples from a repair on the leg made with a different alloy, and 52 samples from classical Mediterranean objects, selected from 800 specimens of ancient leads and ores previously analyzed. Although the article itself is a case study, we discuss here the general data from previous studies presented in the appendix.

As is common in archaeometric lead isotope studies, they presented the isotope data as three ratios to Pb 206. In order to compare analysis with isotope ratios to analysis with isotope fractions, we transformed the ratios back to fractions. The transformation program, its results, and subsequent analysis files are given in Appendix A.2.

Regardless of which set of variables is used, the first step is to present the data as given or a summary thereof. Brill et al. list their ratios to four or five significant figures. One of the problems here is that the first one or two digits are always the same, making it difficult to see the significant variations in the data.

Figure 4 presents selected lines from their table as well as two methods of suppressing the redundant leading digits. The specimens in the table are ordered according to the Pb 208/206 ratio. This ratio always begins with 2.0, with the following three numbers showing all of the variation between samples. In order to allow the similarities and differences between specific specimens to become more apparent, it is better to list only these three digits, using a heading to indicate the magnitude. In this way the important information, which would otherwise be lost in the middle of a large number, is more visible. By erasing what is the same in all numbers, what is different can be more readily seen.

In the same way, all of the Pb 207/206 ratios, except for sample 721, begin with 0.8. The ratio for sample 721 (0.9354) is probably a typographical error, as it is far from all other samples. It probably should be 0.8354, and has been assumed to be such in our analyses. This anomalous value is more apparent with the alternative data presentations. With long strings of numbers an error is easily buried, but when only the significant numbers are listed the error stands out clearly.

In addition to allowing the trends of the data in the table to be visible and improving the chance that errors or anomalies will be noticed, the shortened data would facilitate computer entry of the data for additional analyses by other researchers who may wish to make use of them. Statistical analyses will not be affected by use of the shortened data, and any good statistical package will allow one to convert data in one form into another form automatically.

Another way to present the data for a single variable is a histogram. Examples for both fractions and ratios are shown in Figure 5. Outliers and major typing errors are readily apparent in a histogram. The shape of each distribution is also visible. They are mostly bell-shaped except that the distribution of Pb 207/206 is highly skewed. This distributional asymmetry is a frequent result of taking ratios and is undesirable since most statistical procedures assume that the distribution of values is symmetric, if not roughly normal or Gaussian (standard bell shape).

Figure 4.
Lead isotope data

A. From Table 1 of Brill, Barnes, and Murphy (1981)

Sample Number	Pb 208/206	Pb 207/206	Pb 204/206
616	2.0676	0.8339	0.05312
617	2.0687	0.8341	0.05318
618	2.0693	0.8341	0.05311
1202 (horse leg)	2.0714	0.8348	0.05321
733	2.0717	0.8354	0.05330
673 (horse leg)	2.0719	0.8357	0.05309
729	2.0716	0.8359	0.05335
721	2.0746	0.9354	0.05330
664 (horse body)	2.0754	0.8377	0.05334
1010	2.0825	0.8407	0.05360

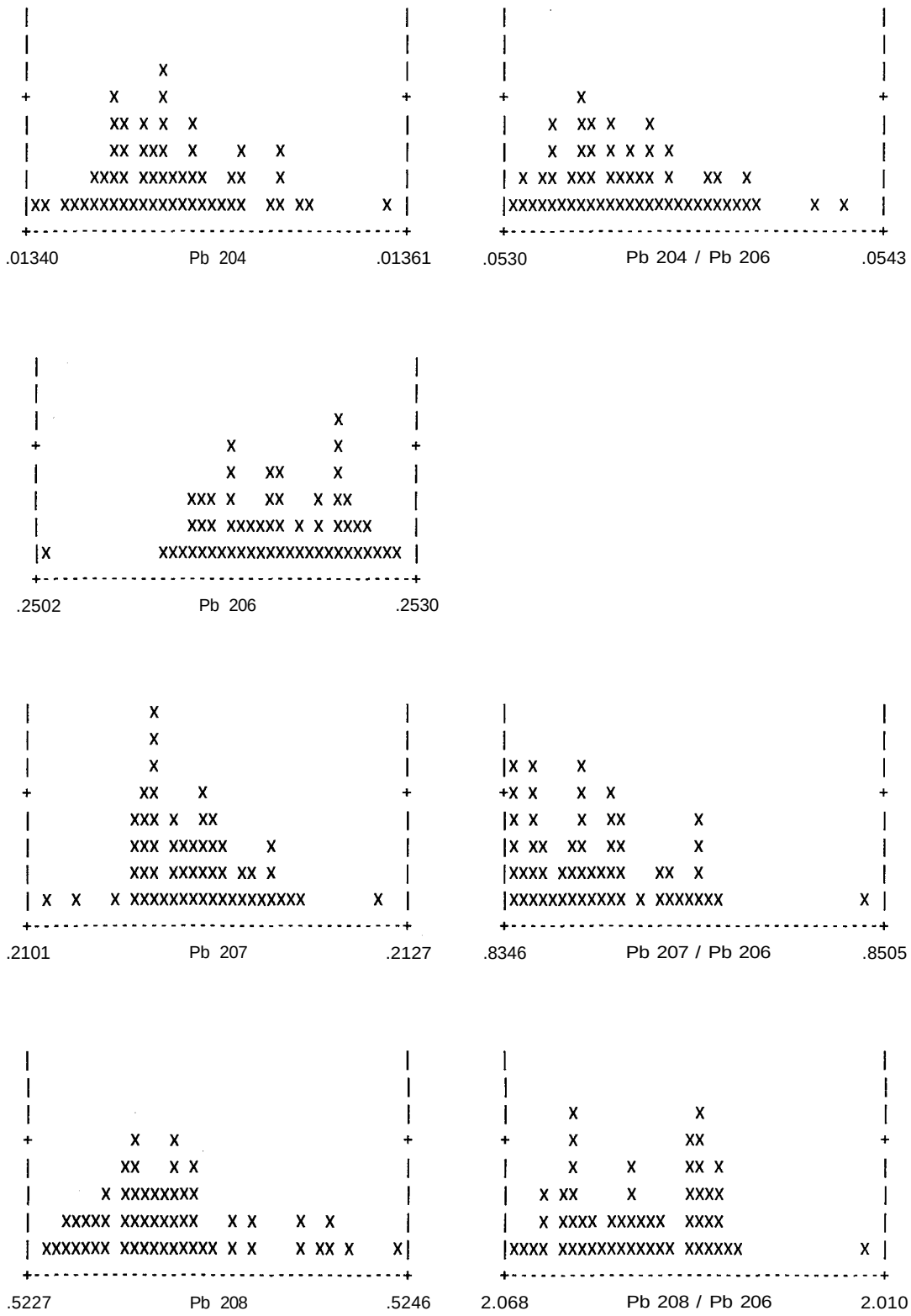
B. With Leading Figures Suppressed

Sample Number	Pb 208/206	Pb 207/206	Pb 204/206
616	2.0676	0.8339	0.05312
617	687	341	18
618	693	341	11
1202 (horse leg)	714	348	21
733	717	354	30
673 (horse leg)	719	357	09
729	716	359	35
721	746	.9354	30
664 (horse body)	754	.8377	34
1010	825	407	60

C. With Constant Subtracted and Decimal Point Shifted

Sample Number	Pb 208/206 -2.06,x10e4	Pb 207/206 -083,x10e4	Pb 204/206 -0.053,x10e5
616	76	39	12
617	87	41	18
618	93	41	11
1202 (horse leg)	114	48	21
733	117	54	30
673 (horse leg)	119	57	9
729	116	59	35
721	146	1054	30
664 (horse body)	154	77	34
1010	225	107	60

Figure 5.
Histograms of lead isotope fractions and ratios



Variable Interrelationships

The next step is to consider the variables together. Their interrelationships can be summarized with correlation coefficients as given in Figure 6.

Figure 6.
Lead isotope correlations
(using Brill, Barnes,
and Murphy 1981:34-36)

	204	206	207	208	204/206	207/206	208/206
204	1.0						
206	-.65	1.0					
207	.73	-.56	1.0				
208	-.03	-.57	-.36	1.0			
204/206	.95	-.86	.73	.22	1.0		
207/206	.76	-.89	.88	.13	.90	1.0	
208/206	.50	-.96	.33	.77	.74	.73	1.0

This table shows that individual isotope fractions have correlations ranging, in absolute magnitude, from .03 to .73, while the three ratios to Pb 206 have correlations of .73 to .90. Since 204, 207, and 208 all have a substantial negative correlation to 206, dividing by this common factor introduces the variation of 206 into the other numbers and thereby increases their correlation. This is the second disadvantage of using ratios rather than fractions or percentages of the total.

The relationship between pairs of variables can be more fully presented with scatter plots of one variable against another. Examples are given in Figure 7, which plots 208 directly against 207, and Figure 8, which plots the corresponding ratios 208/206 and 207/206. The specimens are spread farther apart in the first plot than in the second, because of the lower correlation. The plot of 204 and 208, which are nearly uncorrelated, is even better in this respect.

Figure 8 replicates the plot given by Brill et al. and is the traditional presentation of lead isotope data. It is well established that nearly all samples fall along a rising diagonal. The classification of individual samples is done according to their location along this diagonal, while the distance from the diagonal is effectively ignored. This means that essentially one piece of information, a linear combination of Pb 207/206 and Pb 208/206, is actually being used and that the samples might just as well be plotted along this single dimension rather than in a misleading two-dimensional plot.

There is another plotting method even better suited to fraction data. Three numbers adding to 1, with two independent pieces of information, are often plotted in a ternary diagram. Four numbers adding to 1 require a quaternary plot, with three dimensions for the three pieces of information. With lead isotopes, the 204 fraction is nearly 0 with a range of variation (in this data set) about 10% of the others. Therefore the true three-dimensional plot is well approximated by a ternary plot of Pb 206, 207, and 208. This does not imply that differences in Pb 204 are any less important as a discriminating variable than the other three. In Figure 9 the Pb 204 values are split into five groups, which are represented by the following set of symbols: . - + * #. Using a plotter the symbols could be replaced by circles whose diameter, area, or density represent the amount of Pb 204. Note that the correlation of 208 with the tertiary combination of 206 and 207 forming the x axis is lower than with either alone.

Figure 7.
Lead isotope fractions:
208 vs 207

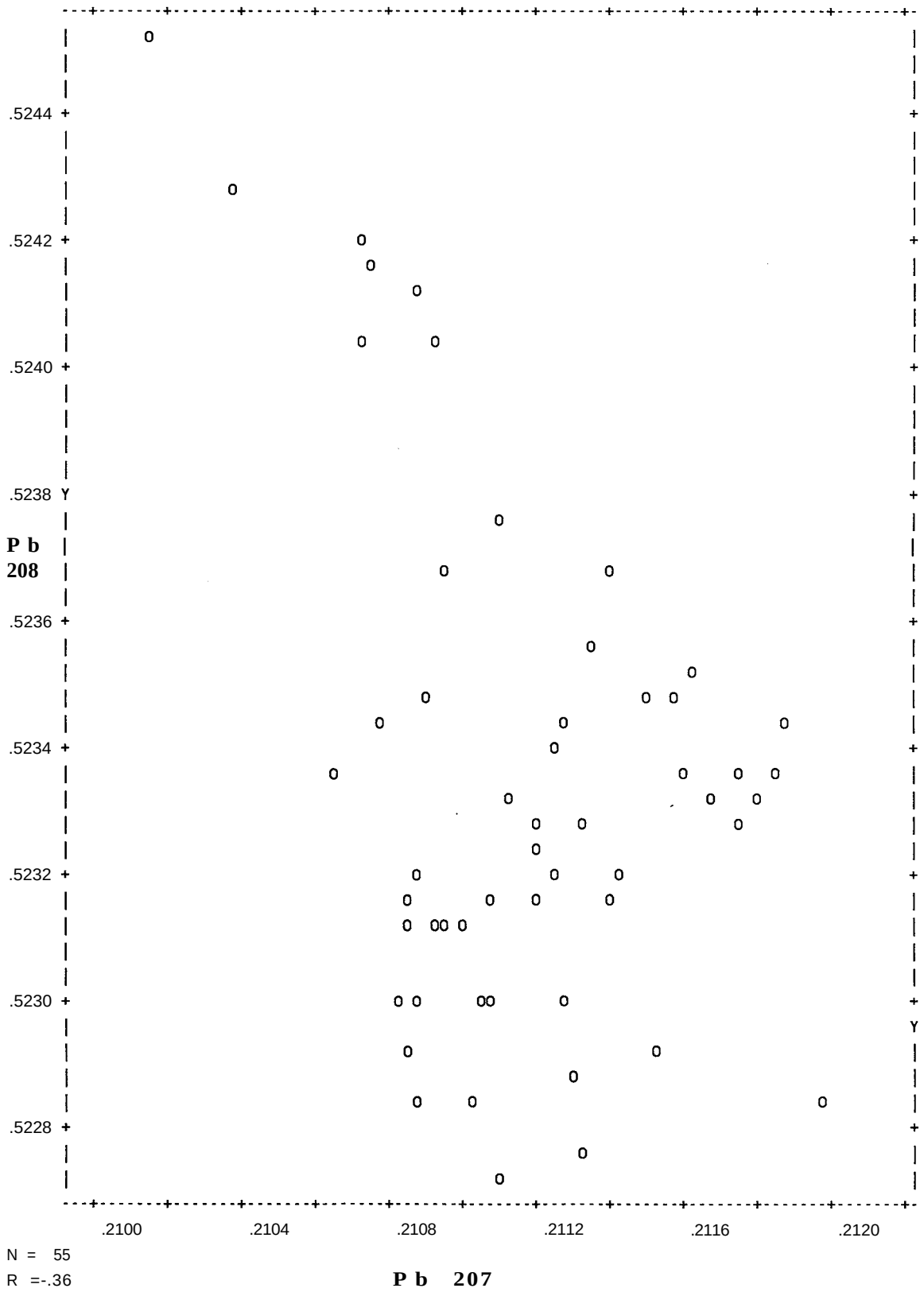


Figure 8.
Lead isotope ratios:
208/206 vs 207/206

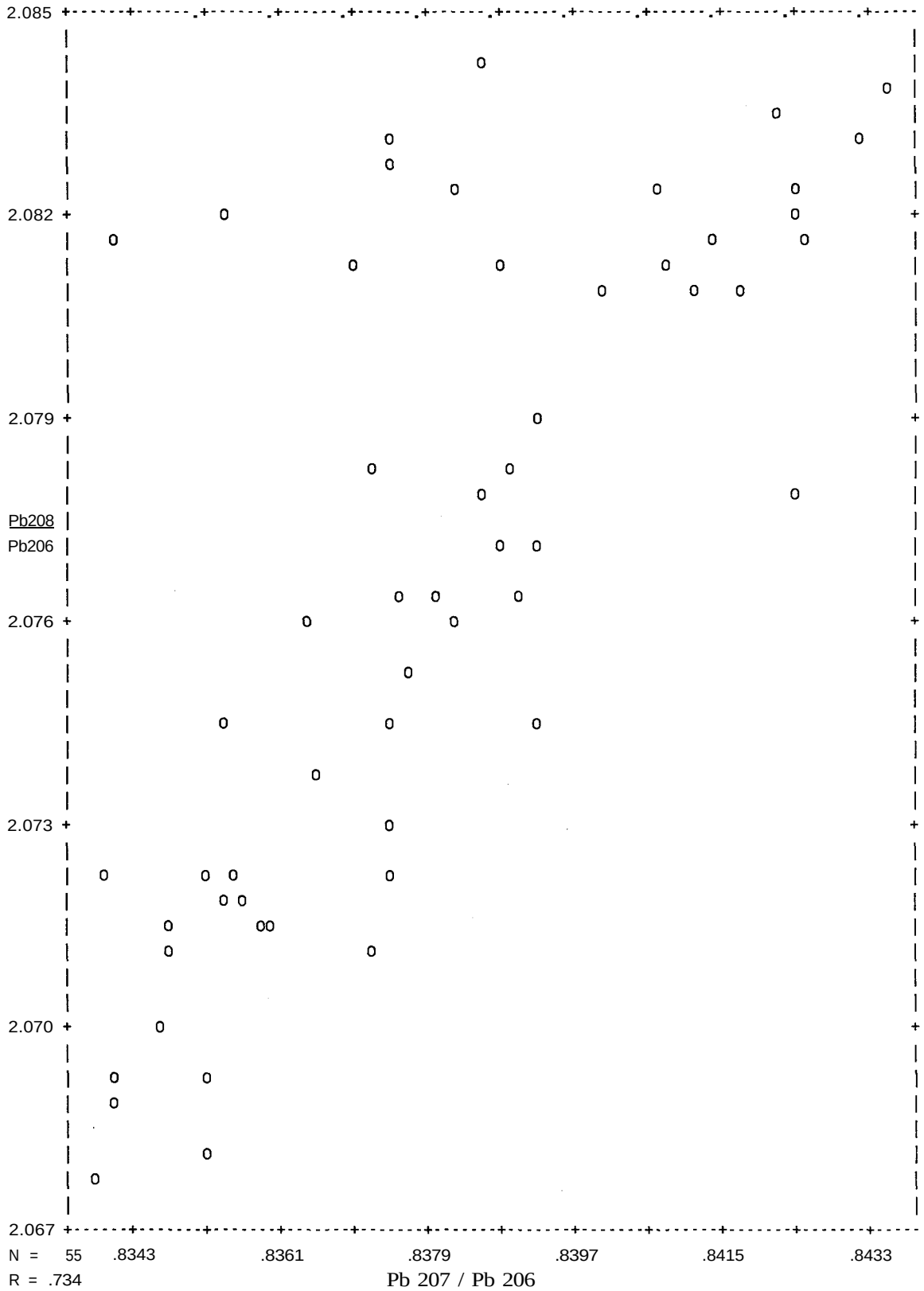
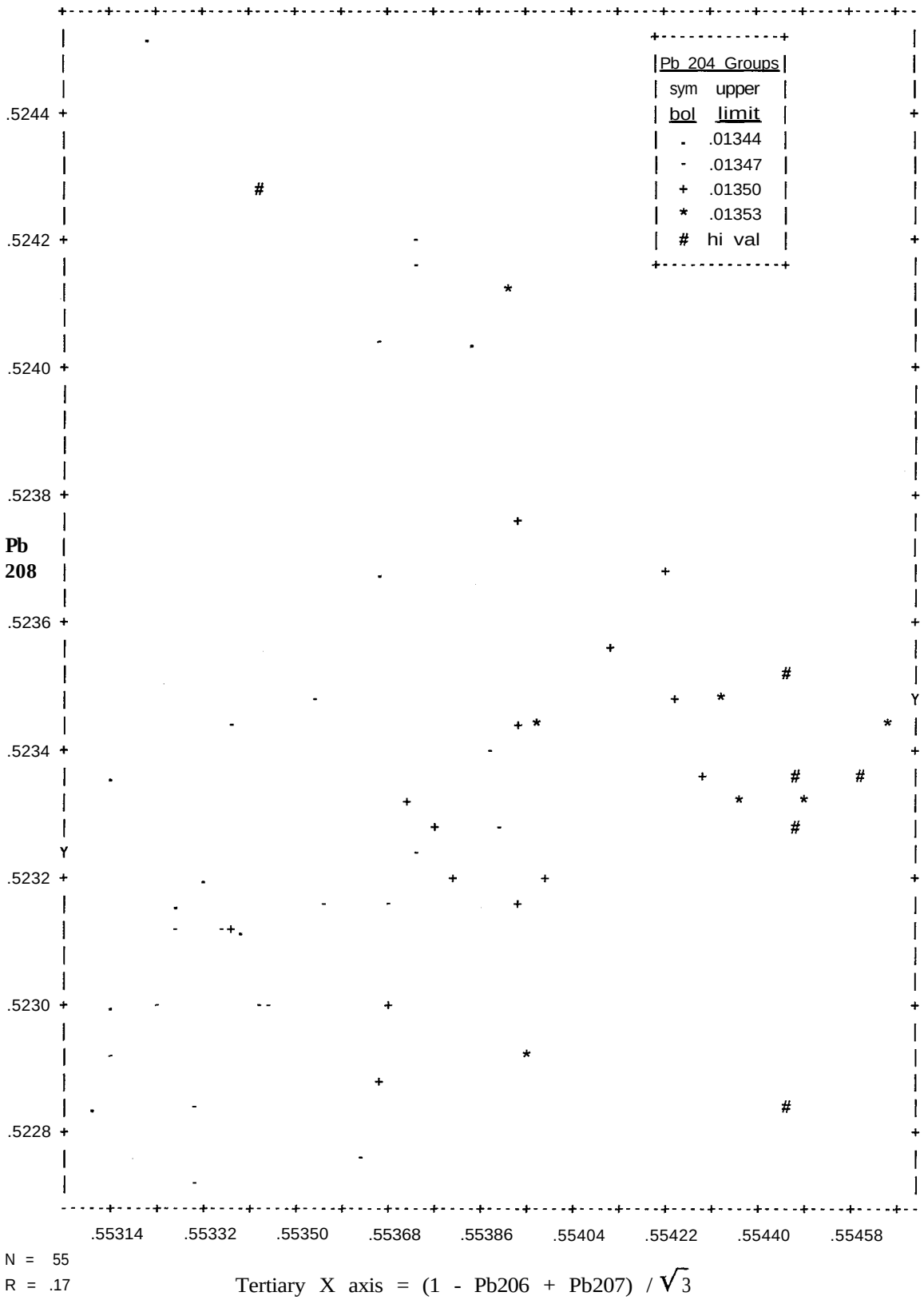


Figure 9.
Lead isotope fractions in a multi-symbol ternary plot approximating a quaternary plot



In geological studies where the purpose is to date the formation of ore bodies, a one-dimensional presentation of the samples is what is needed. In archaeology and art history, spreading the samples out in at least two, or even three dimensions, rather than effectively only one as in the traditional plot, makes better use of the information available. It is more likely to show the true relationship between samples and to visually separate what are actually different groups.

Grouping Structure

A third stage of lead isotope analysis is to examine the multivariate grouping structure of the data. This can be done with either a cluster or discriminant analysis, depending upon whether or not the data have already been divided into groups. In addition to these standard grouping methods, one could do either ordination analysis or multidimensional scaling, techniques that order specimens in a continuum. Points can be labeled by time period or region, and the data examined to see if groups are apparent, or if the data actually form a gradient.

These alternative methods of analyzing lead isotope data will be investigated and presented more fully in a separate project. Before applying the statistical techniques outlined above, we will first separate data for ore sources from data for objects. The first step in a provenance study must be to compare the within-source correlation structure to between-source correlation structures. This information about the nature of lead isotopes in ore sources will allow us to determine what statistical procedures are most likely to succeed in provenance determinations for objects.

Statistical Tests of Significance

In a few instances the degree to which analytical results support or confirm the hypothesis under question is immediately obvious with only a visual scan of the data. Usually, however, the situation is not so clear, so it is good practice to routinely do simple statistical computations that will clarify results.

For an example of a simple statistical test useful for general studies of art material compositions we will use data on East Asian pigments presented by John Winter (1981). The paper addresses two research questions about the occurrence of lead chloride versus lead carbonate in East Asian paintings: Does lead carbonate actually occur in Chinese paintings, and is the situation for Japanese paintings the same as for Chinese ones? Samples of lead white pigments from 45 paintings—13 Chinese, 29 Japanese, and 3 Korean—were analyzed by X-ray diffraction. The results were discussed in light of historical evidence concerning lead white pigments.

Unlike many other reports, all data and final results are given, not just representative information. The author performed appropriate descriptive statistics, with some of the data being counted and clearly summarized in table form (p. 93). In addition, the author does not stop with pigment identifications, but goes on to include relevant historical discussions.

In order to definitely say that the frequency of lead carbonate occurrences is significantly different between Japan and China, a contingency table analysis should be done using the Fisher exact test. This

test gives the exact probability of an observed degree of apparent association in a 2 x 2 table given the hypothesis of no actual association. The more familiar Chi-square test is applicable to any size table and is easier to calculate but gives only an approximate probability value and requires more samples to be really accurate. Figure 10 shows how the data presented in Winter's Tables 3 and 4 can be cross-tabulated for this test. The Fisher exact test shows that the probability of getting such a skewed pattern if composition and provenance were paired at random is less than 1%. In this case the data clearly show that there is a significant association. It is extremely unlikely that such a degree of association between place and white pigment composition is the result of random chance.

Figure 10.
Frequency of white pigment occurrences (data from Winter 1981:92-93)

Painting Origin	Lead Carbonate	Other	Total
Japan	8	21	29
China and Korea	15	1	16
Total	23	22	45

Chapter 3

Deterioration Studies

Organization

The numbers of studies pertaining to the deterioration of art materials of each type in each journal are given in Figure 11. The category of deterioration identification procedures includes the development of methods for measuring the condition of an art object. Again, case studies are limited to one or a small set of objects, without regional or chronological generalizations. General studies focus on a regional or chronological group. Environmental effects include light, heat, moisture, and reactive chemical exposures.

Figure 11.
Frequency of art deterioration studies

	Journal				
Type	SC	JC	TB	PP	ALL
Identification Procedures	6	2	0	0	8
Case Studies	8	0	0	2	10
General Studies	10	2	1	3	16
Environmental Effects	5	2	0	2	9
All Deterioration Studies	29	6	1	7	43

Deterioration studies may be carried out on either real art objects or simulated art materials. Studies of real objects generally focus on what deterioration has actually taken place. Studies with simulated materials look at what changes might take place under various conditions. Statistical analyses are generally easier with the latter, since objects may be generated as needed and manipulated according to a predetermined experimental design. The corresponding statistical techniques are generally aimed at estimating the size of effects of different factors and determining whether the effects are, in some sense, significantly different from zero.

In general, regression accounts for the effect of continuous variables such as time and light intensity. Analysis of variance measures and tests discrete factors such as type of dye. However, continuous variables may be applied at a small number of discrete levels, so the two types of analyses overlap more than might be at first apparent, and can actually be considered as variations of the same basic procedure.

An important step in experimental design and analysis is to identify what are the experimental units and how many there are. The number of experimental units is usually denoted by "n" or "N." The next step, which applies to nearly all deterioration studies, is to identify what are the repeated measurements made on each experimental unit. The third step is to identify what are the treatments and conditions applied to experimental units as a group and how many replicates there are for each combination. The key to correct analysis of deterioration studies lies in correctly answering these questions.

For example, consider two different experiments. In the first, 20 plaster casts are made and 10 are randomly selected to be placed in a

humid environment while the other 10 are kept in a dry environment for the same period of time. The strength of all 20 is measured at the end of the period. In the second experiment, 10 casts are kept first in a dry environment and then in a wet environment and the strength measured at the end of each period. In both cases, there are 20 measurements, but N is 20 in the first experiment, and 10 in the second, with two repeated measurements. These two designs are different and must be analyzed differently. The first would be analyzed with a two-group t-test whereas the second would need a paired t-test. (T-tests are a special case of analysis of variance applicable when there are only one or two groups.)

The typical hypothesis in significance testing is that some effect is equal to 0, although in practical terms the hypothesis is actually that the effect is small enough to ignore. When N gets very large, all effects that are not absolutely 0 begin to look "statistically significant" even though of no scientific or practical importance. The question is, are they different enough from 0? In order to determine the appropriate sample size for a given study, the researcher needs to decide what range of effects is effectively 0, and for the purposes of the study, what deserves attention.

The concepts of regression, analysis of variance, experimental units, replicates, and repeated measurements, and treatments (alone and in combination) will be discussed in more detail in the context of specific examples in the following sections.

Deterioration: Identification Procedures

Color photographs exposed to light fade and suffer changes in overall color balance and may even fade in dark storage if kept at normal temperatures and humidity. Therefore a method is needed to accurately monitor changes taking place in a photographic collection over time. Wilhelm (1981) gives a method to measure color and optical density over time using an electronic color densitometer either directly on the print or indirectly using a fading monitor. When certain limits are reached the print will no longer be considered suitable for exhibition.

Wilhelm's data illustrate the usefulness of an analysis of variance. In his Table III (p. 57) he presents density readings from three different color densitometers, with three different color filters (red, green, and blue), for five film types. From a visual examination of these data he concludes that different color densitometers may give significantly different readings for the same print samples. A statement such as this warrants the performance of a statistical test to check for the significance of densitometer differences, after effects of the other variables have been removed.

Regarding each film as the experimental unit, we performed a 3 x 3 repeated measures analysis of variance (RANOVA) with densitometer and filter type as the repeated measures factors giving nine measurements on each film. The results in Appendix A.3 show that the densitometers are in fact significantly different and that this effect is consistent across all three color filters. Densitometer 1 gave a higher density reading for all 15 film-filter combinations than densitometer 3.

Regarding each densitometer as the experimental unit, we also did a 5 x 3 RANOVA, which shows that the different films also give

significantly different readings. This indicates that not all of the measurement differences are due to densitometer differences and that the effect of differing film must also be taken into account.

Johnston-Feller, Feller, Bailie, and Curran (1984)

They investigated whether the degree of fading of pigments can be quantitatively measured in terms of the changes in concentration of the colored materials. Paint films using alizarin lake as a colorant with titanium dioxide white were exposed to radiant energy in a xenon-arc Fade-ometer filtered to approximate solar radiation through window glass in the near ultraviolet and visible spectral regions. Spectrophotometric reflectance measurements were taken before and during exposure, and computer color-monitoring calculations were made of the percentages of pigments remaining after exposures for various lengths of time. Munsell notation and CIE color-difference calculations were used to develop curves to show relationships of pigment concentration change to fading of pigments.

For each individual pigment-covered plate, they fit a straight line to the logarithm of the relative concentration as a function of exposure and obtained the decay constant. This is a type of repeated measures analysis in which the repeated measures are replaced with a single summary measure. These were then tabulated along with the initial percent of alizarin lake and amount of titanium dioxide.

A possible alternative is to fit an exponential curve to the raw percentage data instead of a straight line to the log percentage. Sometimes after a certain number of hours they have an anomaly in the curve. Linear fits to log-transformed data tend to be more easily thrown off by such things than exponential fits to raw data. To investigate whether the apparent anomalies are due to biphasic decay (different components decaying at different rates) a nonlinear analysis would be almost mandatory. The regression analysis that they did is correct and appropriate to their research problem, but they might find it helpful to do a nonlinear rather than, or in addition to, the linear fit.

Deterioration: Case Studies

The case studies in the journals surveyed are all concerned with documenting the deterioration that has taken place in a particular object or small set of related objects and determining the specific cause. For the type of data presented in these studies, statistical analysis is not applicable. The main question is the within-object choice of samples, as previously discussed for composition case studies.

Deterioration: General Studies

Simunkova, Brothankova-Bucifalova, and Zelinger (1985) researched the influence of various types of cobalt blue pigments on the drying process of linseed oil. They used five different pigments, with each at four different concentrations in linseed oil. These mixtures were spread on glass plates

and then weighed at a series of time intervals as they dried. Drying curves were plotted as was the change in weight against time. The time to maximum dryness was determined visually from these curves. (After the volatile components left, there was apparently some absorption of water so that weight increased.) Two samples were measured for each combination.

This study has both a categorical factor (the pigment type) and a continuous factor (the concentration in weight percent) potentially influencing the outcome variable (the number of days to maximum dryness). An analysis that combines both of these types of factors in a combination of analysis of variance and regression is called analysis of covariance (ANCOVA), with the continuous factors called covariates. The result of such an analysis (Appendix A.4) shows that both pigment and concentration have a highly significant effect on number of days to maximum dryness.

In this study, one can simply look at the data, as did the authors, and be fairly confident that there was a differing effect for different types of pigments. This is because of the relatively low variation between replicates and consistency across different concentrations. With an ANCOVA we can formalize this procedure and make a statistical test of the effect. Simultaneously, we can both estimate and test the concentration effect, which is much harder to do by eye.

The research question in the paper by E. René de la Rie (1982) is essentially an ANOVA question. He sought to determine the effect of various pigments on the fluorescence and yellowing of dried linseed oil used in oil paintings. An analysis of variance would be the basic method to determine whether the pigments have an effect.

His research design was to measure fluorescence spectra of oil paints: lead white after a daylight exposure, then after the daylight plus four dark periods; vermilion after two different daylight exposures, then after a third daylight plus a dark period; lead white and cobalt violet after four different daylight periods; and lead white-vermilion mixture after one daylight exposure. Three pigments on an actual oil painting were also measured before and after removal of the varnish layer.

All data are presented as fluorescence spectra with intensity, wave number, and wavelength. No statistical analysis was performed; instead it was determined from visual observation that the spectra look different for different pigments.

One could measure the significance of differences quantitatively and include a repeated measures test for the cases in which pigments were repeatedly measured after the same series of exposures. The research design could be improved by using the same series of light and dark exposures for each pigment. Then the effect of exposure patterns could be assessed for all pigment types combined.

Deterioration: Environmental Effects

Fading and Dye Mordants

Patricia Cox Crews (1982) sought to determine whether the type of mordant used makes a difference in dye fading. She used 17 natural yellow dyes derived from American plant materials in combination with 5 commonly used mordants to dye worsted wool flannel samples. Two wool

samples with each dye and mordant combination were exposed to light and tested for color change after cumulative exposures of 5, 10, 20, 40, and 80 AATCC Fading Units by instrumental methods and at the end of the total exposure of 80 Units by a visual examination by three trained observers.

Because the design for this study is both appropriate for the goal and clearly presented, it is worth explaining clearly how the resulting data should be analyzed. Her experimental design is described in Figure 12.

Let us ignore, for the moment, the fact that there are repeated measurements on each sample and pretend that there is only one number for each sample. Then we would do an analysis of variance with two grouping factors—dye and mordant. The 170 pieces of information would be divided into what are called "degrees of freedom" (DF) as specified in Figure 12.

Figure 12.
Design and analysis of dye/mordant fading experiment

A. Number of Measurements

	5	mordants
×	17	dyes
=	85	experimental conditions
×	2	replicates
=	170	experimental units or samples
×	5	measurements at different times on each sample
=	850	instrumental DeltaE measurements
		or:
×	3	visual assessments by different people
=	510	visual assessments

B. Analysis of Variance Table

Effect	DF	SS	MS	F
Overall mean of all 170 samples	1	-	-	-
Dye effect	16	-	-	-
Mordant effect	4	-	-	-
Dye-by-mordant interaction	64	-	-	-
Replicate variation (error)	85	-	-	
Total for experiment	170	-		

The missing values in the other columns (-) cannot be filled in because the data are not presently available to us. Associated with each line of this "analysis of variance table" would be a "sum of squares" that reflects the size of the corresponding effect on fading. Just as the degrees of freedom of the first five lines partition and account for (add up to) the total degrees of freedom (number of samples), the sum of squares for the same five lines partition and add up to the sum of the squares of the 170 numbers. Each sum of squares would be divided by its corresponding degrees of freedom to get the "mean square" (MS) or variance. The variance for each of the first four lines would be divided by the replicate or error variance (fifth line). This ratio of variances is known as "F." It measures how much variation is introduced into the data by the effect under consideration in relation to the amount of variation due to random experimental effects. If an effect is actually null, or nonexistent, then F should be about 1. The probability of getting an F value of a given size for given degrees of freedom for effect and error terms (the "p value") can

either be computed by known formulas or looked up in standard tables that have the results of such computations. P values less than .05 or .01 are usually called "statistically significant."

It is desirable to include replicate experimental units in an experiment, where "replicates" refers to multiple experimental units given the same combination of treatments. Suppose that there are no replicates or that the replicate values have been replaced by their mean. Then there would be no proper error term in the analysis of variance table since this is entirely due to the replicates. We would then have to assume that there is no dye-by-mordant interaction. This would make the expected mean square for interaction equal to the now unavailable mean square for the replicates, and we must use this as the divisor for calculating F for the main dye and mordant effects. (Having assumed that this effect is 0, we can no longer test whether it is otherwise.)

If the assumption of no interaction effect is true, then the resulting F value will be about the same as if it were calculated using the replicate variance as the divisor. The corresponding probability or p value will be higher due to the lower denominator degrees of freedom, but noticeably so only if the study has far fewer experimental units than this one. On the other hand, if the interaction effect is significant, the replacement denominator will be noticeably larger, making the F values smaller than they really ought to be and the corresponding F or variance-ratio test less powerful (less likely to discover true differences) than it would be if the replicate variance were available and used as the denominator.

Now let us consider the fact that each experimental unit is measured three times (visually) or five times (instrumentally). In respect to dye and mordant effects, what difference does this make to the analysis? The answer is, none at all! The multiple measurements must be summarized by one number, usually but not necessarily the mean, and the analysis of the grouping factors carried out exactly as before. In other words, one should not summarize across replicates but must summarize across repeated measurements for the purpose of analyzing factors applied to independent experimental units (there is a multivariate approach to repeated measures using a modified form of multivariate analysis of variance [MANOVA] but this technique is beyond the scope of this review).

When, for instance, one takes three repeated readings with an instrument one right after another, it is standard practice to immediately reduce the three readings to their mean or middle value before beginning analysis. When the repeated readings are separated by days instead of seconds, the principle is no different.

There are two possible purposes for repeated measurements. First, one may be trying to reduce error and especially avoid blunders as would be noticeable if one of three readings were way off from the other two. If there is otherwise no expectation that the three readings should be different and no interest in any possible order effects in taking readings, then the individual readings are not needed. Second, measuring the effect of time or an associated variable such as exposure may be a primary goal of the study. In this case, the individual repeated measurements must be recorded and analyzed. But this analysis is separate from the analysis of the grouping factors.

If an experiment only has repeated measures factors, also known as "trial factors," then it can usually be analyzed as if the trial factors were grouping factors by including the experimental unit as a grouping factor and by using the "experimental unit by trial factor" interaction mean square as the denominator for the F tests. If an experiment has both grouping and trial factors, then one should either use a program such as BMDP2V that knows how to keep these two types of factors separate or else seek out an experienced statistician to make the necessary but difficult adjustments to the output of standard multifactorial analysis of variance programs.

In light of the above, Crews made two major errors in her analysis of the DeltaE data. First, she replaced the replicate values by their mean, resulting in the problems described above, including the inability to test for dye-by-mordant interactions. Second, she did not summarize across the five time measurements in her analysis of dye and mordant effects, but treated time as a grouping factor. The result is that she used an incorrect error term with an inflated number of degrees of freedom for all her tests and thought she was testing for dye-by-mordant interactions when she could not. In addition, she did no analysis of the visual assessments beyond a side-by-side comparison with the last DeltaE.

The data she presented in the paper are the mean across two replicates of the last DeltaE after 80 units of exposure and the mean of three visual estimates by different observers of Lightfastness and Gray Scale, which are estimates of color change in comparison to standards. We analyzed all three sets of data by ANOVA with mordant and dye as grouping factors and the mordant-dye interaction as the error term. These results are given in full in Appendix A.5. In all three cases, mordant effects are highly significant while dye effects are not. Crews claimed that dye is also significant, but, as explained above, we cannot consider this valid with the data as given. As also explained above, the result might be different if we had the replicate data.

It is not surprising that the three measures give nearly the same result. The correlation coefficients are .75 for DeltaE to Lightfastness, .68 for DeltaE to Gray Scale, and .90 for Lightfastness to Gray Scale. The two visual measurements are essentially redundant.

We are puzzled at her errors because she acknowledges the help of a statistician. Did she miscommunicate her design? Was he not familiar with repeated measures analysis? Did she misunderstand his directions? Was their statistical software inadequate? Statistical consulting seems to be a difficult enterprise, with miscommunications common both ways. We hope that this technical report will help conservation researchers to be more successful at obtaining statistical advice and assistance useful to their particular problems.

As is typical of many experimental studies in conservation, all samples came from the same type of material from one manufacturer, and probably from the same bolt of wool. For making the comparison she made, this is desirable since it eliminates wool differences as a factor. On the other hand, for generalizing the results to the universe of types of wool, the number of experimental units is effectively 1. This means that while we can *assume* that the results are true for other wools, we have no real *information* about the interaction between different wools and dye, mordant,

and exposure effects. Since she was doing replicates, she might have considered using two types of wool to get some evidence as to differences between wools and whether the same mordants have the same effects on dye fading.

Her Figures 1 and 2, which show the mean color change for each mordant-dye combination grouped first by dye and then by mordant, would be improved if both dye and mordant were sorted by their mean values instead of being haphazardly arranged. An alternative is presented in Appendix A.5, which has a histogram for each mordant, sorted by decreasing mean color change.

Finally, we return to the analysis of the repeated measurements. The five measurements at various exposure levels could be analyzed for linear trend and curvature (quadratic effect). For completeness, cubic and quartic effects, corresponding to double and triple levels in the data, could also be included in the analysis. (This is done automatically by BMDP2V). This analysis could be done with exposure measured either linearly or logarithmically (as implicitly done by Crews's choice of exposures). Alternatively, a nonlinear rate constant could be fit to each set of five measures.

Each single degree-of-freedom measure of exposure is analyzed exactly like the summary across exposures. The line labeled "mean" in the analysis of variance table in the Appendix would be relabeled "exposure." The line labeled "dye" would be relabeled "dye-by-exposure interaction." This process is illustrated in our analysis for the next subsection.

If the visual assessments of individual observers were available, it would be possible to see whether there were any consistent differences between observers. The measure of observer variability would have $3 - 1 = 2$ degrees of freedom. The analysis of variance table would again be the same as before except that all degrees of freedom would be multiplied by 2.

Fading and Light Filtration

Bowman and Reagan (1983) focused on another aspect of fading: whether removing infrared and ultraviolet light rays reduces the textile dye fading known to be induced by various types of lamps currently used in museums. Several 5 x 8.25 cm specimens of bleached cotton cloth were dyed with either turmeric, madder, or indigo, which were chosen to cover a broad range of colors, dying procedures, and lightfastness. The specimens of each type (at least six, but number not specified) were assigned to one of three lamp types (incandescent, fluorescent, tungsten halogen quartz), which were either left bare or covered with the appropriate filter or filters. Color change from the initial state after four exposure times was measured by reflectance readings with K/S values (percentage reflectance at wavelength of maximum absorption read from the spectral reflectance curves, proportional to dye concentrations). The objective was to determine how significant fading is under each of the six different lighting conditions and what the interaction effects are for each lamp-filter-dye type combination.

Using numbers read from their three plots (one for each pigment) of the effect of light exposure on K/S value (pp. 41-42), we did a RANOVA on all the data and on each dye separately (Appendix A.6). There is no replication, so we had to use the highest-order interaction term for the error term. As expected from looking at the plots, there is a highly significant dye effect. Both the light effect and interaction between dye and

light are significant. An examination of the plots and analysis of variance for each dye indicates that this is due to abnormally small changes in indigo after 100 and 200 hours of exposure to fluorescent light. There is a significant filter effect and a linear trend across exposures that is consistent for each dye.

Bowman and Reagan performed an analysis of variance with Duncan's Multiple Range test to determine significant differences in color loss attributed to the lamp-filter systems evaluated. They reported the results of these analyses in their text but omitted the analysis of variance table that would allow us to determine their exact model. If, however, they included time as a grouping factor along with dye, lamp type, and filtration this would be an error similar to that made by Crews since time represents repeated measures on a single experimental unit.

Linen Canvas Strength

Hackney and Hedley (1981) studied whether the weakening of linen canvas can be avoided or slowed down by shading, enclosure in a sealed case, and/or impregnation with a bees wax/resin lining mixture and determined acidity effects on canvas strength.

Linen canvas samples that had been aged naturally for 24 years were arranged on three different boards. One board had only impregnated samples. Tensile strength was measured for 30-40 yarns from each exposure condition, with means and standard deviations computed. PH was measured for cold water extracts. Their results are summarized in Figure 13.

Canvas patches kept in the dark are consistently stronger than those exposed to light. Both enclosure and waxing increase the strength of unwaxed, exposed patches, but add nothing in combination. PH, which has a correlation of .75 with strength, has essentially the same pattern, except that the effect of waxing is less than that of enclosure. Analysis of variance (Appendix A.7) confirms the significance of these results.

Figure 13.
Tensile strength and pH of linen canvas threads

Measure	Wax	Shade	Open	Enclosed
strength	bare	light	1.2	2.2
		dark	1.8	2.6
	waxed	light	2.2	2.0
		dark	2.5	2.3
pH	bare	light	4.0	5.5
		dark	4.3	5.7
	waxed	light	4.8	4.9
		dark	5.1	5.2

Hackney and Hedley spent four pages comparing each pair of canvases differing by a single factor with a t-test of the individual thread strengths. At best, this procedure determines that the mean thread strength in the two particular pieces of canvas is different. Even here, there is the problem that the experimental unit is a piece of canvas and not individual threads. The true degrees of freedom for the t-test are

probably less than they claim due to correlation between neighboring threads.

Paint Film Yellowing

Levison (1985) studied the yellowing and bleaching of paint films. His research problem was to determine whether exposure to daylight will bleach out dark-induced yellowing discoloration in paint films and whether the degree of darkening and its susceptibility to bleaching is a function of the age of the paint film or a function of the previous darkening-bleaching cycles the object has undergone.

The experimental design was to use three drying oils in white pigments with a variety of paint mediums. An initial Yellowness Index (YI) was measured, then the test panels were exposed to four cycles of dark and light. YI was measured after each stage, and the net change from initial was computed.

Because a series of tests were made on the same specimens after various cycles of exposure to light and darkness, a repeated measures analysis is needed. Levison appears to have done the mental equivalent of a paired t-test comparing the initial and final results. The months of yellowing and bleaching and corresponding mean YI values calculated from his Table 5 are in Figure 14.

Figure 14.
Mean yellowness index after alternating light and dark

Exposure Type	Exposure interval (months)								
light	4.6	2	4.2	2	4.1	6	4.6	25	4.3
dark		10		6		10		13	

We can reject the hypothesis that all four dark means are the same (Appendix A.8). It appears that a longer dark period leads to more yellowing. There is no linear trend across the five bleached measurements. The higher order variations, however, are significant. The changes from one measurement to the next seem too consistent. The initial drop from 4.6 to 4.2 results from a decrease in 24.5 of 32 samples (no change is counted as .5) and the increase from 4.1 to 4.6, an increase in 26 out of 32 samples. These are significant even with a simple binomial sign test (same as asking, "What is the probability of 26 or more heads or tails in 32 coin tosses?"). Instrumental drift might be an explanation. It might also explain some of the dark variation.

The paint films tested should be broken down into appropriate subgroups for analysis. Levison discusses various subgroups in his conclusions but it is not clear enough which samples he includes in which subgroup to proceed with this.

Ozone-induced Fading

In a study on the fading of traditional natural colorants due to atmospheric ozone, Whitmore, Cass, and Druzik (1986) examine the rate at which various natural colorants deteriorate upon exposure to ozone. A total of 16 organic materials derived from plants and insects, commonly used as colorants prior to the development of synthetic coloring agents, were tested. Each was exposed for 12 weeks in the absence of light to an ozone

level equivalent to heavy smog. Fading from the original color level was measured instrumentally by diffuse reflectance spectra.

Their classification of colorants (Table 4, p. 121) as very reactive, reactive, possibly reactive, and unreactive seems arbitrary and contributes little information beyond what is in the plots (p. 120), which show little evidence of discrete groupings. We would certainly draw the lines between colorant groups in different places if forced to make groupings at all.

There are two important pieces of information in these data: the maximum fading for each colorant and the rate at which fading occurs, or runs towards the maximum. Some materials may fade more slowly than others upon exposure to ozone, but keep on fading longer (saffron, for example). Thus, the relative order of dyes with respect to amount of fading may change with increasing exposure. Each curve could be fit by a hyperbola or negative exponential to get a projected maximum fading and fading rate.

The authors appear to assume that the observed fading differences can be generalized to other samples, but no evidence is presented here for that, as no replicate analyses are included. The implicit assumption is that if the study were repeated with the same colorants, the curves would look the same, and therefore the observed curve differences are the result of real dye differences and these differences will persist if we repeat the experiment. But that assumption is not backed up with data. It is possible that these differences could also be due to a high variability in testing procedures and results. Therefore, it would be better to analyze fewer types of colorants if necessary to do at least a few replicates. Repeating only a selected subset of the colorants would give us information that would help us judge the reliability and repeatability of the results for the other colorants.

The interpretation of their results is also hindered by the lack of control samples. The data as given do not demonstrate that the fading observed was caused by ozone. A few samples prepared with all conditions the same as for the others but with no ozone introduced would be appropriate for comparison.

Chapter 4

Conservation Treatments and Materials

Organization

As with composition and deterioration papers, conservation papers can be subdivided by study type (identification procedures, case studies, and general studies), as listed in Figure 15. To vary the presentation, we have chosen to organize this chapter according to the study steps of experimental design, data presentation, and statistical analysis, and subdivisions thereof. These correspond to the variables used in the statistical survey presented in Chapter 5.

Figure 15.
Frequency of conservation studies

	Journal				
Type	SC	JC	TB	PP	ALL
Identification Procedures	15	15	3	5	38
Case Studies	17	12	10	31	70
General Studies	19	12	1	7	39
All Conservation Studies	51	39	14	43	147

Experimental Design

Number of Research Conditions or Treatments

These are usually very clearly stated in the conservation literature. However, there are two types of exceptions. In the first, the experimental design and analytical tests are not clearly stated in the text of the paper, but have to be inferred by the reader from the results table (Simunkova, Smejkalova, and Zelinger 1983). In the second, a different number of research conditions is given in different parts of the paper. For example, in a case study involving experiments to assess the potential for using microwave radiation to disinfest wool fabrics (Reagan 1982), seven exposure times are given in the methods section (p. 21), while in the results section (p. 25), eight exposure times are listed in both the text and in the table.

Number of Replicates and Repeated Measures

The key concept for correctly reporting the number of replicates analyzed is understanding the difference between replicate samples and repeated measures on a single sample. When the repeated measures are separated in time, this is fairly clear. Examples are fading experiments in which one sample is measured several times.

Less clear are situations where the measurements on a given experimental unit are separated in space rather than time. In agricultural research, these are called split-plot experiments. As an example, consider an experiment involving the analysis of a new adhesive mixture being considered as a conservation material. If many samples are taken from one

batch of the adhesive preparation, the multiple analyses are repeated measures of that batch. If many batches of adhesive are prepared and one sample is taken from each batch, the multiple analyses are replicates. Repeated measurement of one variable is also different from simultaneous measurement of multiple variables such as fading, strength, and weight.

The primary experimental unit for a study is (or should be) that type of unit which forms the class of entities to which one wants to apply the results of the study. Confusion as to this point leads to confusion about whether particular measures are replicates or split-plot repeats. If all samples in an adhesive study are obtained from one batch, and they are analyzed as replicates, then the results strictly apply to that batch only. If the samples are analyzed as split-plot repeats, then they jointly characterize the batch as a whole and the summary results can be extended to whatever class of batches this batch is considered to be a part of. However, with only one batch, we would have no internal evidence as to how representative the one batch is for the entire class. We must either make an outright assumption, such as "all batches are the same and have the same internal variability," or have some prior evidence about the variability of batches. If we want our results to apply to the class of batches, rather than to the class of aliquots from a single batch, then we gain more information about the population of batches if we take samples from different batches.

Confusion over the difference between taking many measurements and samples from one object and obtaining replicates from multiple objects is one problem area in conservation research experimental designs. There are many cases where the intent of an experiment is clearly to obtain results generalizable to a class of objects, yet the "replicate" samples are actually repeated measures on one object. In other cases only one sample is analyzed.

In either case, the number of measurements may be unstated, vague, or contradictory. It is important for correct interpretation of results that the reader be able to discern these aspects of the experimental design. A composite (made-up) example typical of many reports is, "Samples were taken from two rolls of wool fabric." In other cases, the exact number of both objects and samples is given, but it is left unclear as to how many received each of the particular treatments being tested.

Nosek described the conservation of an eleventh-century lead paten excavated in Krakow. The corrosion products were identified by X-ray diffraction and energy-dispersive X-ray fluorescence. The experiment description says only that "Spectral analysis was performed twice, both before and after conservation treatment" (1985:20). We would like to know whether the analyses are based on measurement of one or multiple areas and whether the before and after measurements were performed at the same location(s).

Some other examples of numerically vague statements are the following: "...thin sections of gypsum were prepared..." (Skoulikidis and Beloyannis 1984); "...test fabrics were cut to the proper dimensions with an NAEF die..." (Block 1982); and "...a number of wrought iron nails and pieces of cast iron were immersed..." (Gilberg and Seely 1982).

A clear statement of the number of replicate samples is given in Branchick, Keyes, and Tahk (1982), which reports on experiments concerning the bleaching of naturally aged paper by artificial and natural light.

Their Table I (p. 33) indicates the exact number of samples that received each particular treatment.

Sampling Design

The sampling design for a study should allow generalizations to be made at the level intended. It is important to include at least two replicates at this level so that variability can be assessed. If repeated measures are used, it is best if each treatment is applied to subsamples of each object. If samples are selected randomly, the method of randomization should be described. If not, the rationale for selection should be given.

Generalizations

Samples should be drawn from the population that the researcher wants to generalize to. The situation in conservation research is somewhat different from that often encountered in chemistry or physics. Except for minor impurities, reagent-grade chemicals will be the same from batch to batch. But most art materials receiving a conservation treatment are inherently heterogeneous, variably structured, mixtures. For example, due to differences in raw material, processing, and aging, all "paper" is not the same. Therefore if all samples in a study are derived from one roll of paper or one bolt of cloth, the experiment itself gives no idea of how well one can generalize to other rolls of paper or other bolts of cloth.

For example, Barger, Krishnaswamy, and Messier (1982) studied the effect of four tarnish removal methods on one simulated nineteenth-century gilded daguerreotype. Each method was applied to one strip with a fifth left untreated as a control. The surface of each strip was tested for overall fading by measuring the total reflectance of highlight and shadow regions. Changes in image particle size and distribution and average number of particles per given area were determined by scanning electron microscopy. There is no description of how similar the strips were before treatment, nor how treatments were assigned to strips.

On the basis of this single sample it is difficult to recommend one of these treatments over the others. The authors say they also performed the same experiment on an ungilded daguerreotype but did not report the results because it was "less representative of nineteenth-century daguerreotypes." However, the results from the second sample would have given some indication of the effect of gilding and the consistency of the relative performance of the treatments. If all nineteenth-century daguerreotypes are actually gilded, then their second simulated sample could have been also, with each gilded daguerreotype divided into five strips, giving two replicates of each treatment method.

Variability

Because many conservation studies are intended to allow a conservation treatment method or material to be recommended or condemned for use on art objects, it is important that such experiments include some replication to assess the potential variability in treatment results, and to safeguard against errors that may lead to a "fluke" result. However, we encountered many studies with an effective sample size of one. Studies using real art objects can often be designed to allow replication, and studies with simu-

lated art materials can always do so. Often, analyzing or treating fewer types of objects, but including replicates, would greatly improve the reliability of the study.

The importance of replication is shown by Phillips (1984). To answer the question of whether an acrylic precipitation consolidant can work well for strengthening some leathers, one sample of two types of leather objects and two samples each of two other types were treated. The two nineteenth-century calfskin replicates gave very different results. Because of that variability in results, he concludes that the treatment cannot be recommended now but does deserve further study.

Repeated Measures

In a split-plot experiment where different treatments are applied to different parts of an object, it is best if each treatment is applied to some part of each object. This allows all treatments to be compared on the same group of objects. It also allows the use of standard computer programs for the analysis of repeated measures. Such programs require a complete design without holes or missing values.

Barger, Giri, White, Ginell, and Preusser (1984) studied two coatings and a control treatment on 17 nineteenth-century daguerreotypes. Each treatment was applied to one-third of each daguerreotype for a complete repeated measures design. The only question remaining is how the treatment assignments were made within each daguerreotype.

Clement (1983) tested nine hydrogen peroxide bleaching treatments of stained and discolored paper (including three controls) for blistering side-effects. Seven expendable lithographs were cut into pieces and distributed among various treatments. Since the smallest lithograph yielded 25 sections, the best design would have been to apply each of the nine treatments to at least two pieces from each of the seven lithographs. The actual design has several holes.

Parrent (1985) tested three methods of stabilizing water-logged wood with sucrose along with a control of no treatment. Three archaeological woods were split into four pieces to receive each of the four treatments. Several others were kept intact and treated as a whole with one of the treatments. The comparison of treatments in such a partial repeated measures design is more difficult than if it had been entirely repeated measures or entirely separate artifacts.

In biological studies repeated measurements may be taken on a single rat. Alternatively, rat litters may be used as experimental units with the individual rats within a litter receiving one of several treatments as "split-plots," which are similar in both genetics and developmental environment. In either design the problem is that a rat may die during the experiment. In art conservation studies this problem does not arise, making it easier to do complete repeated measures experiments.

Randomization

Random sampling and random assignment of treatments to samples, to be differentiated from haphazard methods, requires that a method of randomization be followed. The possibilities include physical randomiza-

tions such as coin flipping and drawing well-mixed tags out of a container, random number tables, and computer-generated pseudorandom numbers. At least until these become standard practice, the method used should be specified when reporting the experiment.

None of the studies reviewed in the conservation literature that reported using random sampling described the method of randomization that was used. For example, Block (1982) mentions that treated and untreated samples were chosen for aging "at random," but says nothing further.

Selection Rationale

If samples are selected or assigned in a particular structured manner for specific reasons, these should be stated, as they may affect interpretation of test results. In the conservation literature reviewed, the method or reasons for sample selection are rarely given.

One class of rationales is based on spatial relationships. Peacock (1983) examines whether deacidification agents successfully used in paper conservation can also reduce the rate of deterioration of a cellulose fiber textile (flax linen) during accelerated aging tests. Three deacidification agents were tested, each with two application methods. Each agent-application combination was applied to ten samples. The assignment was done so that "Within each group of ten specimens no two samples had warp or weft threads in common. Therefore, samples were structurally independent of one another" (pp. 9-10).

Another reason for particular selections is to cover a spectrum of possibilities. Alessandrini, Dassu, Bugini, and Formica (1984) wanted to determine the composition of materials used to construct the Roman period chapel of St. Aquilino in Milan in order to design a conservation program. They took 38 samples representing all previous restorations and different states of preservation. A large range of analytical tests were performed to identify the mineralogy, chemical composition, total soluble salts, morphological and structural characteristics, and physical characteristics. These data are used to deduce the state of preservation, mechanism of decay, and best choice of restoration procedures.

Randomization can be combined with such structured designs. In the flax aging example, each group of 10 carefully selected specimens could have been randomly assigned to a particular treatment. Similarly, randomization could have been used to select samples within a restoration-preservation class of the cathedral materials.

Data Organization

Based on our review of data presentations published in the conservation literature, we make the following suggestions for improvement over current organization methods.

Tables

The numbers in a table should include an appropriate number of actually significant figures (digits). The table should be labeled so that it is clear what each number represents: one measurement or the mean or other

summary of many, and if the latter, how many. The tables should be organized to make the primary comparisons most clear. When there are multiple related tables, their organization should be made as consistent as possible.

Significant Figures

It is hard to make a mechanical rule, but the general guideline is: Think about which numbers in your data set are meaningful, only report numbers that mean something, and consider your purpose in presenting them and to what they are being compared.

For example, if a measuring instrument or process or computer printout gives a number with several digits but you know the uncertainty is more than 1%, present only three or even just two digits depending upon whether the number is above or below the nearest power of 10. The number 93, which is below 100, has an uncertainty in this case of at least .9 and could be written with only two digits. On the other hand, 11.3, which is just above 10, has an uncertainty of about .1 and should be written with three digits. The exact choice is partly a matter of personal judgment and the particular situation.

Similarly, when calculating a number such as a percentage based on the ratio of two counts, only present the digits that are real, even though you can carry the calculation out indefinitely. Each count can be considered to have an uncertainty of plus or minus one-half count. If the denominator of the ratio is a count of 25, then the uncertainty in the numerator becomes an uncertainty of plus or minus 2% (100% x .5/25) in the resulting percentage. The uncertainty in the denominator usually makes the uncertainty of the result even higher. Tacking on decimal fractions of a percent would be inappropriate and misleading.

Often, in order to make comparisons clear, one can profitably round off before the uncertain digit with little loss of real information, even though this is contrary to most peoples' initial instincts. If after rounding, all numbers have the same trailing zeros, these can be deleted and the units appropriately adjusted in the table title or legend. Similarly, if all numbers have the same leading digits, these can be subtracted from everything in order to make the differences more obvious, and an appropriate explanation given.

Labeling

Properly organized raw data tables are necessary for analysis. Summary result tables are usually necessary to present the results of analyses. It should be clear to the reader of a table what each number represents, whether a raw datum, transformation thereof, or summary. In any case, the units should be clear and in the case of summaries it should be clear what is being summarized, including how many. Without picking any particular examples from the literature reviewed, we note that many tables were unnecessarily obscure.

Organization

The structure of a table is part of its information content and therefore deserves some thought to improve its communicative potential. For

example, the comparison of two numbers is easier if they are juxtaposed vertically rather than horizontally (side by side). If a table is going to present comparisons in both directions, then, other things being equal, the primary comparison should be in the vertical direction. Furthermore, the decimal points should then be lined up.

It is fairly common for data to have multiple categorizations. For human readers it is easier if multiple lines in the same category are labeled on the first line and successive lines left blank. When a table is intended to be used as input to a statistical program, then other rules apply. Every line should be completely labeled and the data otherwise organized as required by the particular program, possibly with header removed and categories coded.

Figure 16. Comparison of tables for humans and for computers

For humans:

Species	Mordant	Fading
llama	alum	3.0
	iron	2.7
sheep	alum	2.8
	iron	2.6

For computer:

la	3.0
li	2.7
s a	2.8
s i	2.6

Data Availability

There are several reasons to publish data resulting from an experiment. Doing so allows the reader to:

1. get a feel for the nature of real data of the particular type presented;
2. verify and extend the statistical analysis;
3. ask different questions of the same data;
4. combine results across studies;
5. experiment with new methods of statistical analysis;
6. use the material as a teaching example.

These are all legitimate scientific purposes that can only advance our knowledge and techniques.

It is our opinion that raw data tables, if not included in the paper, should at least be made available upon submission to journals. That practice would allow reviewers to check data analyses and judge validity of interpretations; the data should then be made available to journal readers who may wish to follow up a particular study. It is the explicit policy of *Science* that papers are accepted for publication with the understanding "that any materials necessary to verify the conclusions of the experiments reported will be made available to other investigators under appropriate conditions" (*Science,* editors 1987).

In most cases in the conservation work we have reviewed, the raw data table would take up only a page or less. If too voluminous to publish, copies of data sheets should be made available on request, preferably from a central depository. It is sometimes claimed that because scientists "own" their data, they have a right to keep it "secret." However, we feel that once a scientist makes a public claim about experimental results, the reader has a reasonable right to see the supporting evidence if it is easily retrievable.

Plots

Plots are an alternate means of presenting both raw data and summary results. As with tables, plots should be self-explanatory if at all possible, rather than requiring the reader to search the text in order to be able to interpret them. The most appropriate occasion for presenting data in plots rather than tables is when there are at least two ordered variables. Whether the ordering is over time, space, or quantity, the relationship of such variables is easier to see with plots rather than tables.

Plots should be clearly labeled so that it is immediately apparent what the data points represent, single samples or means. There should be a key that identifies the meaning of different plot symbols and any other unusual characteristics of the plot. Organizational methods applicable to tables are also applicable here, such as ordering and labeling variables consistently in a series of comparative plots.

The comments in the paragraph above are based on actual examples where the suggestions given were not followed. In addition, there are apparent inaccuracies or inconsistencies where, for instance, samples are shown as beginning with less than 100% of full strength at time zero. If there is an explanation other than an inaccuracy in drawing the plot, it should be reported.

Statistical Analysis

Descriptive Statistics

There are three ways in which the use of descriptive statistics in conservation research can be improved. First, give the number of items averaged (already discussed in the section on Tables). Second, correctly calculate, use, and differentiate between standard deviations and standard errors. Third, use descriptive statistics in many situations where they are currently absent.

Standard Deviations and Standard Errors

The standard deviation of a batch of numbers is a measure of how far apart or how variable the numbers are. In particular, it is the root mean square deviation from the mean or average. In other words, subtract the average from each number, square the difference, find the average of these deviations, and then take the square root. If we are interested in the standard deviation of a population but only have a subset or sample of the population, then we cannot calculate the standard deviation of the population directly but must estimate it from the sample. To get an unbiased estimate, modify the formula by dividing by N-1 instead of N when calculating the mean squared deviation. The direct calculation is called the population standard deviation, whereas the indirect estimate is called the sample standard deviation. In most experimental studies the latter is what should be used, although it only makes a noticeable difference with small sample numbers.

Just as individual measurements differ from object to object, summary measures (statistics) differ from collection of objects to collection of objects. In testing hypotheses about summary measures or statistics we need to know how much they would vary if we were to repeat the entire experiment. If we do not want to repeat an experiment several times to

actually calculate a standard deviation for the statistic, we must look for an easier method. It turns out that we can estimate what the standard deviation of the summary measures would be by dividing the standard deviation of the individual measurements by a factor that is typically proportional to the square root of N. Such an estimate of the standard deviation of a summary measure from the standard deviation of the measurements it is summarizing is called a standard error.

Some authors (Pearlstein, Cabelli, King, and Indictor 1982; Nelson, King, Indictor, and Cabelli 1982) have reported a "standard deviation at the 90% level of confidence." However, neither the statistician writing this technical report nor another professional statistician we consulted have ever heard or read this particular phrase. It is thus unclear what they meant.

Potential Uses of Descriptive Statistics

Descriptive statistics, particularly the computation of totals and percentages, could allow additional use of the data already collected in the course of conservation case studies and general studies of real materials. These summaries might identify overall trends and thus aid in conservation treatment decisions.

Two papers containing case studies of wallpaper conservation included sample forms that were used to collect data on wallpaper conditions and treatments at various historical sites. Clapp (1981) describes the types of information routinely collected from wallpaper samples at Winterthur, with a brief discussion of the reasons for collecting each type of information. Gilmore (1981) also presents a form used for collecting information about wallpapers. Compared with the more arduous tasks of identifying what is important to record, creating the forms, and collecting the data, combining the results of all the forms into a descriptive summary would be a relatively simple procedure. This effort might shed light on both conservation and art historical problems. For example, Clapp lists criteria that distinguish Oriental from non-Oriental wallpapers. It would be interesting to see a count of how many samples in actual practice fall into each category and a discussion of the effectiveness of the criteria.

Many of the variables appearing in such forms will be categorical (color, origin, material) rather than numerical (size, weight). Individual categorical variables are summarized by counting the number of objects falling into each category (a frequency distribution). The relationship between categorical variables is examined by cross-tabulation tables. For example, is there any relationship between the type of paper used to make wallpaper a century ago and its condition today? This descriptive summary procedure is appropriate and potentially useful to any class of objects or treatments.

Most case studies carefully describe all of the materials that were used, the number of objects conserved, and sometimes the cost of the materials involved. An additional descriptive statistic appropriate for conservation case studies is the approximate time required to complete the recommended or described treatment. Such a time estimate can give other conservators the information they need to decide if they can or should proceed with that treatment themselves. Times are rarely given, but a

good example is found in the paper by Thomas McClintok (1981). This case study describes how a one-color wallpaper was conserved *in situ,* with some different treatment problems encountered than appear with patterned wallpaper. The area treated (240 square feet) took 93 hours, with 24 hours for surface cleaning, 42 hours for mending, and 26 hours for filling and in-painting.

Estimation

Estimation methods more complicated than the calculation of simple descriptive statistics have rarely been used in art conservation research. All regression analyses encountered used linear methods. In many cases, nonlinear fitting would be more appropriate. Extrapolation of a linear approximation to a curvilinear relationship may produce dubious results. Mathematical linearization tricks that allow linear regression usually introduce other problems. In any case, these compromises are no longer necessary since computer programs for doing nonlinear regression are now easily available.

In the paper by Skoulikidis and Beloyannis (1984) on reconversion of gypsum into calcite, the function they call parabolic is specifically exponential. It is only parabolic in the general sense of curving either upward or downward but not both, and not in the specific, well-defined sense of being quadratic in the dependent variable. They do not say how the fitting was done, but we can determine that they must have used linear regression. It would have been useful to have some of the data presented to show how many points were fitted, what the average deviation from the curve was, and how much smoothing they did.

Hypothesis Testing

In order to generate an answer to a research question from observed or experimental data, description and estimation often need to be followed by inference or hypothesis testing. After summarizing a group of items and their variability and estimating some aspect of the population of interest and the uncertainty of our estimate, this third step may be needed to connect the results to what we want to know. Although hypothesis testing is a major part of statistical technology, it has been little utilized in the conservation literature.

Two major methods of hypothesis testing are t-tests and analysis of variance (ANOVA) for both grouping and repeated measures factors. Analysis of variance has already been discussed in several specific contexts. This subsection covers the principles of these methods in a more general framework. One of the analyses suggested for specific conservation studies was performed but most of the corresponding papers did not have sufficient data to do so.

Although the computational details of hypothesis tests can be confusing, the basic principles are fairly simple. The four steps, which should become clearer by the end of this subsection after some specific applications and examples are presented, are:

1. Select a "null" hypothesis that is neutral with respect to the effect being studied. This conservative negative hypothesis is the one that is directly tested.
2. Generate from the data a summary measure of the size of that effect as evidenced in the data. This is usually a measure of

deviation from or variability about the neutral condition hypothesized in step 1.

3. Divide the empirical value of this summary measure by an estimate of how large it ought to be if the null hypothesis being tested is true. This estimate depends upon the number of samples, their variability, and the scale of measurement. The purpose of this division is to standardize the summary measure so that one can compare it from experiment to experiment as well as against standard tables. When the null hypothesis is true, values of this standardized ratio (test statistic) near 0 are fairly common whereas values far from 0 are relatively rare.
4. If the observed test statistic is large enough to be very unlikely to have come about if the hypothesis being tested were true, then reject that hypothesis and entertain an alternative that makes the observed value more probable.

In such tests, the probability of getting a ratio at least as far from 0 as that observed is called the p value. It is standard practice, but not mandatory, to reject the hypothesis when the p value is less than either .05 (1 chance in 20) or .01 (1 chance in 100), depending upon how conservative one wants to be.

The appropriate null hypothesis to test may depend on the current knowledge and practice in the particular area being investigated. If there is no known treatment for a particular condition, then the null hypothesis is that a proposed treatment has no effect, i.e., that it is effectively the same as doing nothing at all and no better or worse. If there is an established treatment known to be at least partially effective, then the null hypothesis should probably be that the new treatment has the same effect as the existing treatment, rather than none at all.

The procedure outlined above at first seems a bit backwards: To prove that something is so, we assume that it is not and then show that the negative assumption should be rejected. However, this is a statistical application of the philosophical principle of William of Ockham, which suggests that explanatory entities, in this case "effects," not be multiplied beyond necessity. The specific application to medical and conservation practice is that treatments not be applied unless shown to have sufficient benefit to justify the cost and risk of unwanted side-effects. While it can be overdone and turned into a mechanical ritual, hypothesis testing has become quite useful since its development in this century.

One Group

Given a set of measurements on a sample from a population, we can test the hypothesis that the mean value of the measurement in the population is 0. We divide the sample mean by its standard error to get a t statistic (ratio) whose p value can be calculated under a certain set of assumptions.

This procedure is more general than it seems. To test whether the population mean is any fixed value other than 0, subtract that value from all measurements and do the test as described above. Repeated measures on each sample can be summarized in any fashion desired to get one number per sample, which is then tested. The exact hypothesis being tested depends upon how the repeated measures are summarized. If there are

two measurements per sample and the difference is calculated for each sample, then the hypothesis being tested is that the average difference for all members of the population is 0.

There are other methods of testing hypotheses about a single group, such as the sign test and signed rank test, which have various advantages and disadvantages relative to the t-test. We shall not discuss these further here.

Bomford and Staniforth (1981) studied whether mixtures of beeswax and either Dammar or Ketone-N resin, applied to the back of painting canvases, change the color on the front. They prepared canvases of various thicknesses with various historical grounds. There was at most one replicate of each combination. Each canvas was divided into thirds, each section getting either one of the mixtures or a control treatment. DeltaE, a measure of color change, was then determined for each section.

The 14 canvas-ground combinations are a selection from the universe of possible but realistic prepared painting canvases. The first null hypothesis is that the average effect of the two wax-resin mixtures is the same as the effect of the control treatment. We calculated for each canvas the difference between the mean DeltaE of the two wax-resin sections and the DeltaE for the control section. After assuming that a dash ("-") in their table's column for control treatment DeltaEs means 0 and eliminating an oddball canvas for which the resin-control difference is relatively huge, we got a t statistic of 3.8, which has a p value of less than .01. We thus reject the hypothesis of no difference and conclude that these wax-resin mixtures have a statistically significant effect on increasing the DeltaE measurement.

The second null hypothesis for this experiment is that the two resins are equivalent. To test this, we took the difference for each canvas of the resin DeltaEs and got $t = 2.67$, which has a p value less than .02. Ketone-N causes significantly more color change, on average, than Dammar.

Because the results were not 100% consistent, in that Dammar caused more change in 3 out of the 14 canvases, Bomford and Staniforth said that their results, "do not suggest that one mixture has a greater effect on color or darkening over the other." However, the more careful statistical analysis described above indicates that their experiment is indeed powerful enough to differentiate between the two resins and answer their research question.

Multiple Groups

Given exactly two groups of samples, as defined by some difference in condition or treatment, the usual null hypothesis is that the corresponding population means for some variable are equal. This hypothesis is the same as the hypothesis that the difference between the two means is 0. The observed difference between sample means is divided by its standard error, based on the standard error of the two means, to get a t statistic as with one group. A rank sum test can also be used for testing this hypothesis.

With more than two groups, the usual null hypothesis is still that all group means are equal. However, the test procedure is slightly altered to use variances (mean squares) rather than differences and standard

errors. The observed variance of the group means is divided by the expected variance of the group means, which depends on group number, sizes, and the variance of individual measurements. This ratio is called the F statistic, and the procedure is called analysis of variance.

This type of analysis can be extended to more complicated situations in which samples are grouped by more than one factor. The general model for an analysis of variance is that the observed data is a linear combination or sum of effects of the various factors and their interactions, plus a random residual or error term. This model is similar to the model used in linear regression. The least squares estimate of the effect of each treatment or condition or combination thereof is the mean for all samples subject to that particular treatment, condition, or combination thereof. There is a corresponding null hypothesis as to the effect of each treatment, condition, or combination.

The most thorough analysis of variance encountered in the set of 320 papers reviewed was presented by Wang and Schniewind (1985). Their research was concerned with consolidation of deteriorated wood with soluble resins, and what effects type of soluble thermoplastic resin, molecular weight of the resin, type of solvent, resin concentration, and drying rate of solvent have on improvements in strength and stiffness of the wood. A total of 580 specimens are included in the study, taken from four Douglas Fir foundation piles removed from the ground near the San Francisco waterfront after 70 years of service and deterioration. Among the 145 specimens from each pile, 25 were left untreated as controls while 5 were assigned to each of the 24 treatment combinations resulting from 2 soluble thermoplastic resins, 3 resin concentrations, 2 solvents, and 3 solvent removal (drying) rates. Bending strength and stiffness were calculated from static bending load-deflection curves for each of the 580 samples.

Analysis of covariance was done with wood density as a covariate and treatment and pile as main effects. A 4-way analysis of covariance was done to examine the effect of concentration, drying rate, type of solvent, and molecular weight of Butvar. Two 4-way analyses of variance were done for Butvar, with molecular weight, solvent, pile, and either concentration or drying rate as the main effects. Two 3-way analyses were done for Acryloid resin, with solvent, pile, and either concentration or drying rate as main effects.

Without an analysis of variance table or a more complete description in their text, we cannot be absolutely sure of how they did their analysis. Their inclusion of pile as a factor suggests that all factors were analyzed as grouping factors. An alternative analysis would treat resin, concentration, solvent, and rate as repeated measures or split-plot (split-pile) factors. The authors say (footnote, p. 86), "Since each pile originates from a different tree, and wood properties can be expected to vary from tree to tree, pile was included as a factor in the analysis." This correlation of properties for samples from the same tree or pile is the reason: (a) for using a split-plot design, as they have, and (b) for doing a corresponding analysis that does not assume the lack of such correlations.

Pearlstein, Cabelli, King, and Indictor (1982) measured the effect on paper of rubbing with four different eraser products. One type of paper was aged before and after erasure according to four different protocols.

Folding endurance, tensile strength, and surface pH were measured. Crumbs were removed in half of the samples. Thus, they did a study with three factors—four aging protocols, four eraser types, and two eraser crumb removal methods.

Analysis of variance of all the data would simultaneously examine the effects of all three factors and their interactions. The reason for doing factorial designs is to analyze several factors more efficiently than simply varying one factor at a time, as in the classical scientific experiment. In addition, such designs allow investigation of interaction effects.

Another factorial design appropriate for analysis of variance was used in the research on bond strengths of Lascaux 360 H.V. and BEVA 371 by Katz (1985). Bonding of sized and unsized canvas by each of these two adhesives was tested after activation by one of two methods. A 2x2x2 3-factor analysis of variance for each of the two bond-strength measures (peel and lap/shear) would give a quantitative measure of which main and interaction effects were statistically significant. Again, the analysis of variance would simultaneously test the effect of each of the three factors (adhesive, activation method, and sizing) as well as the interaction between those factors.

Clement (1983) researched which hydrogen peroxide bleaching conditions and pretreatment procedures produce the least amount of blistering on degraded papers. He used seven nineteenth- and twentieth-century lithographs that were cut into small pieces and evenly distributed into groups, each of which received a different treatment (there were a total of nine treatments). Blistering was visually estimated in degrees of damage ranging from 0 to 4, and bleaching was measured by an increase in brightness (reflectance).

An appropriate way to analyze these data would be to first do a repeated measures analysis of the nine treatments. If there were no significant differences between treatment methods, one could then stop. But if results are not the same for all nine treatments, one could then test particular contrasts that stand out as important (equivalent to doing the one-way t-test described above) with the lithograph as the experimental unit of analysis. A contrast is a specific combination of the individual values that highlights a specific effect that one is interested in exploring.

Pia DeSantis (1983) investigated the long-term effect on degraded paper of a strong solution of the protease derived from *Aspergillus saitoi.* She had three factors: two types of paper, which were artificially aged for three days at 100° C; five different treatments (including the control of no treatment), applied to 20 samples each; and post-aging or not for half (10) of the samples for each treatment. All samples were then tested for brightness, fold endurance, and pH.

She analyzed data by doing multiple t-tests, comparing each of the five groups to every other group. Thus she did twenty comparisons where only four independent comparisons are possible. If enough t-tests are done, it is almost certain that one will be significant. A five-group one-way analysis of variance would simultaneously test for differences between the five treatments and lessen the problem of false positives. If treatments and post-aging were applied to samples of each type of paper so that all twenty three-factor combinations actually occurred (this is not clear from the article), then a 2x5x2 three-way ANOVA might be the analysis to begin with.

For most conservation experiments, there is a choice between alternate designs. In the paper/enzyme experiment just discussed, an alternative would have been a split-plot design. Each sheet of paper could have been split in five portions after aging, with each of the portions getting one of the five treatments. These portions could have been split again for post aging. Different designs will result in different amounts of information for each effect for a given amount of experimental effort. The best choice will depend on the details of each situation. One of the reasons to choose a repeated measures design is to get more information about the effects of most interest, even at the expense of less information about other effects.

Chapter 5

Statistical Survey of Conservation Papers

Introduction

In 1986 we reviewed every paper published during the previous five years in four English-language conservation journals. The abbreviations are repeated below:

JC	*Journal of the American Institute for Conservation*
SC	*Studies in Conservation*
TB	*National Gallery Technical Bulletin*
PP	*AIC Preprints*

The JC series began with the Fall 1980 issue and ended with Spring 1985. The others began with the first issue of 1981. A sixth issue of PP, that for 1986, was added when it became available during the review process. This sample of the conservation literature comprised 320 papers.

This chapter presents a statistical analysis of the types of papers published, the types of statistical methods used, and the interrelationships between the two. We expected some changes over time. We were curious about whether or not there are major differences between journals. We expected that there had to be some relationship between the type of study done and the sophistication of statistical analysis.

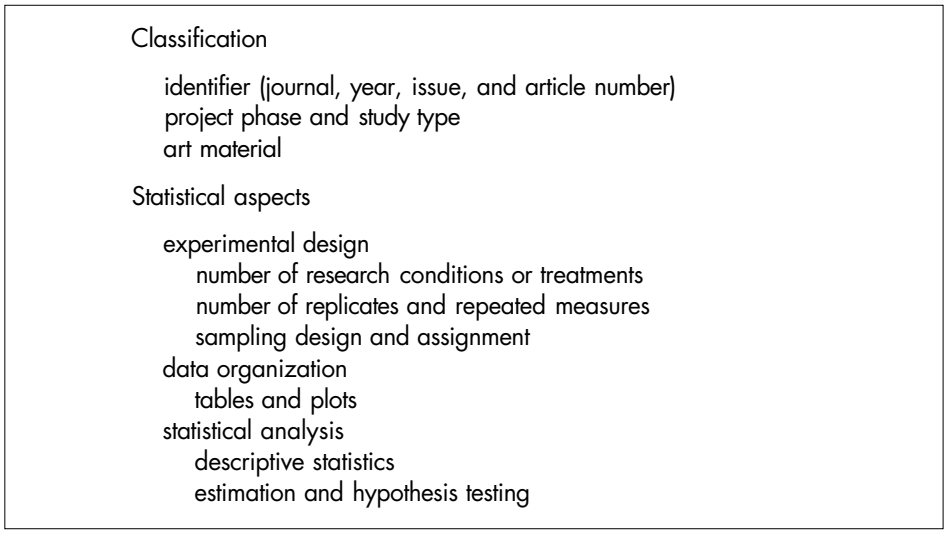
The results of our statistical analysis of the published literature is presented both for its intrinsic interest and as a case example of a thorough statistical analysis. Another reason for presenting this survey is to reveal what statistical methods are currently used in conservation research in order to give readers an idea of which basic statistical concepts to be familiar with in order to be able to fully understand the literature in this field.

Survey Method

Survey Variables

To do a statistical analysis, we must describe and summarize the objects under study with a set of data items that is sufficiently complete to answer our questions. The information evaluated and tabulated for each paper in this study includes the variables listed below. The statistical variables cover an experiment from design to conclusion in the order given. All the variables are listed in Figure 17 and described in full in the following subsections. Data for all papers are listed in Figure 18 at the beginning of the Results section.

Figure 17.
Survey variables in the data file



Classification of Conservation Papers

Identifiers

Identifiers, as used in Figure 18, have four parts indicating journal, year, issue, and sequence number. The first two letters indicate the journal in which the paper appeared, using the abbreviations given above. The first two-digit number refers to the year of publication. The following letter identifies the issue within each year. TB and PP have only one issue per year, so all papers in those journals are labeled "a"; JC has two issues per year, and these are labeled "a" for the Spring issue and "b" for the Fall issue; SC is published four times a year, so these are labeled "a," "b," "c," and "d" for numbers 1, 2, 3, and 4. The final two-digit number of the identifier identifies the numerical order of the paper within a journal issue.

Project Phase

To conserve an art object, one must:

A. determine the composition of the art object or material;
B. consider how it has or might deteriorate;
C. apply conservation materials and methods to remedy current damage or prevent further damage.

Most papers present the result of a study focusing on just one of these three steps or phases of a conservation project. These were easily coded A, B, or C according to their dominant emphasis. Conservation case studies that explicitly covered all three phases were coded C. There were otherwise few ambiguities.

Study Type

Papers were grouped and coded for this variable in one of the following categories:

1. Description of how to carry out a particular procedure or build and use particular equipment.
2. Case study of one or a few real objects.

3. General study of a class of simulated objects.
4. General study of real art objects, includes the general work of one artist.
5. Study of environmental effects on art objects.
6. Essay (literature review, philosophical or museological discussion, or any other paper not presenting primary results).

The difference between types 3 and 4 is the difference between studying the composition, accelerated aging, or consolidant effect on samples from Italian marble quarries and performing equivalent studies on Italian marble statues. The number of papers reporting environmental studies was too small for meaningful statistical analysis so these were assigned to either type 3 or type 4 for the analysis in this chapter.

Art Material

Our original classification of art materials studied is given in Figure 18. Where more than one material was discussed in a paper, the primary material was listed. When a paper focused on a conservation material or treatment, such as adhesives or various chemicals, the art material it was or would be used on is the material type listed. For meaningful statistical analysis with sufficient numbers in each category, we grouped the materials as follows:

metals (iron, copper-based alloys, silver, other metals)
substrates (paper, wallpaper, canvas, textile)
coatings (pigment, varnish, dye, photograph, daguerreotype)
minerals (stone, ceramic, glass, shell)
organics (wood, leather, ivory, reed, lacquer, plastic, moss)
other (analytical method, conservation and exhibition management)

The mineral and organic categories respectively include all nonmetallic inorganics and organics other than those included in the previous categories.

Statistical Aspects of a Study

We coded each paper for its presentation of eight different statistical aspects of the design and analysis of a research study. If an item was missing, we decided whether, given the study's type, design, and purpose, the item was inapplicable or should have been present. If an item was presented in the paper, we judged the clarity and completeness of the presentation. The correctness of statistical analysis procedures, as presented, was also judged. This gave us the following four codes:

1 absent and inapplicable
2 absent but should have been present
3 present but unclear, incomplete (or incorrect)
4 present and clear, complete (and correct)

We realize that these evaluations are sometimes subjective. However, we have attempted to be consistent in the criteria used to perform the evaluations. They are the product of both authors.

Experimental Design

This was divided into three aspects:

1. Number of research conditions or treatments: applicable to papers in which some experimental work has been carried out; the exact procedures followed in preparing and analyzing samples should be clear.
2. Number of replicates and repeated or split-plot measures.
3. Sampling design: the criteria and methods used for sample selection and the assignment of samples to treatments.

Data Organization

The two types of data organization and presentation reviewed are:

4. Tables.
5. Plots.

Statistical Analysis

This is broken down into:

6. Descriptive statistics: totals, percentages, averages or means, and standard deviations or standard errors.
7. Estimation: regression and correlation analysis.
8. Hypothesis testing: t-tests, analysis of variance, and repeated measures analysis.

Multivariate techniques such as cluster analysis and discriminant analysis were never used in the conservation literature, although they might have been, so they are not included here.

Statistical methods that were absent were coded as inapplicable if they were not really necessary to the study as designed; however, there were many studies that could have been designed differently to produce quantitative results suitable for statistical treatment. In these cases, rather than evaluating the design and trying to decide on an alternative, the "inapplicable" code was given. Therefore the large number of studies for which statistical methods were coded "inapplicable" does not actually mean that statistical methods are not valuable, but instead means that many studies in conservation research are not designed to obtain quantitative, testable data.

In some cases, absent analyses were judged "should have been present" when the authors presented conclusions implying that at least the mental equivalent of a statistical analysis was performed. Examples are conclusions claiming "significant differences between treatment results" or "trends in the data."

Survey Data Analysis

Each of the variables coded for this survey was individually summarized using BMDP program 1D. For numerical variables, the program provides the number of valid values, mean, standard deviation, and extreme values. For categorical variables it provides the number of cases (frequency) in

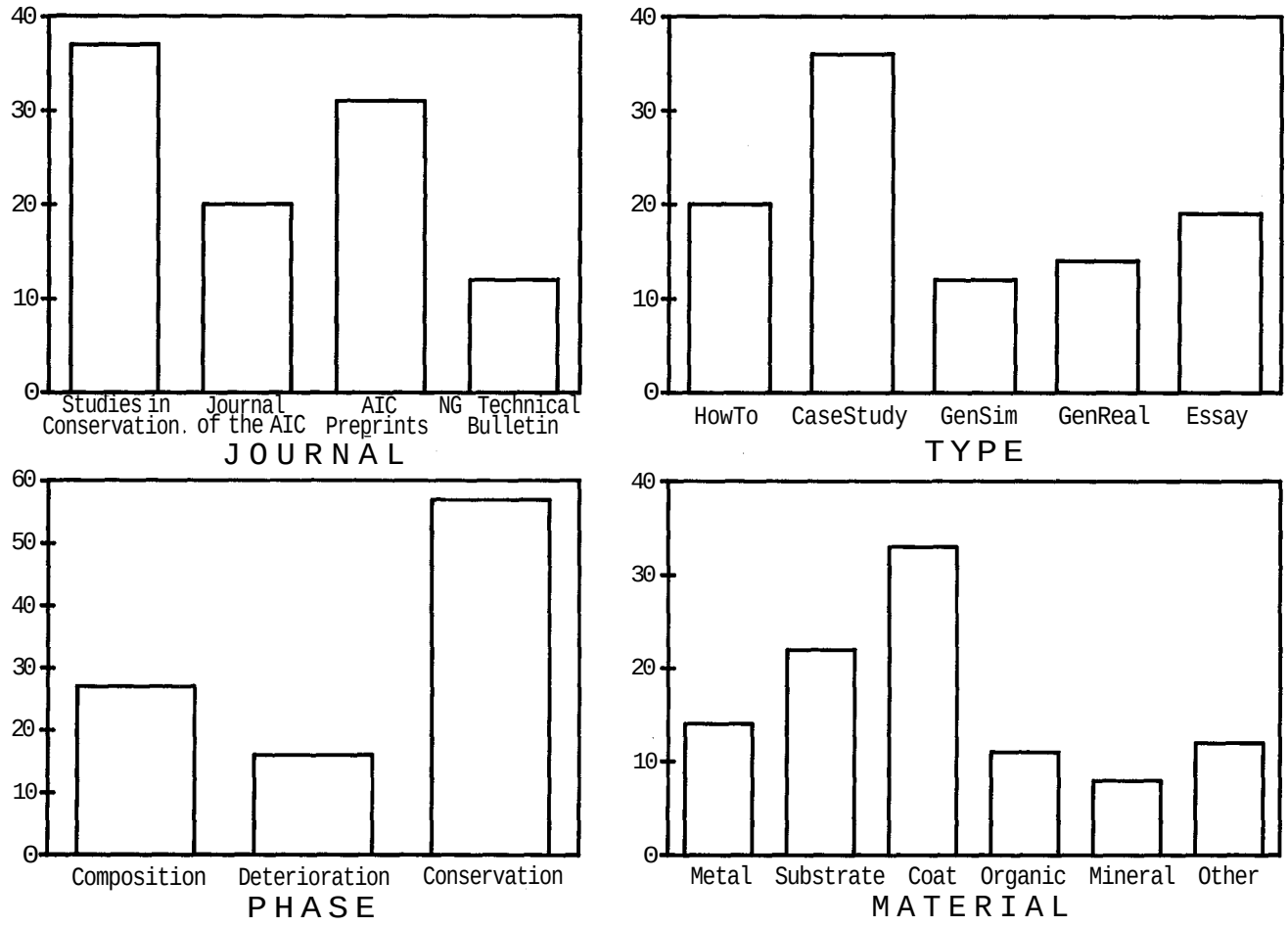
Figure 18.
Data for analysis from survey of 320 art conservation research papers

```
Journal: ABCD = JAIC, Stud.in.Cons., Nat.Gal.Tech.Bul., AIC Preprints
| Year and Issue within year
| |  Article # within issue
| |  | Phase: abc = composition, deterioration, conservation
| |  | |Type: 12346 = how-to, case-study, gen-simulated, gen-real, essay
| |  | ||  Material: see table below
| |  | ||  |  Experimental Design: treatment, reps, sampling
| |  | ||  |  |   Data Presentation: tables, plots
| |  | ||  |  |   |  Statistical Analysis: describe, estimate, test
| |\ | ||  | /|\ /| /|\
A80b01 a4 cv 423 33 411     Codes for Material
A80b02 c1 cv 111 11 111
A80b03 c1 pg 111 11 111     mi  marine iron
A80b04 c1 pp 111 11 111     fe  other iron
A80b05 c6 ot 111 11 111     cu  copper alloy
A81a01 c6 wp 111 11 111     ag  silver
A81a02 a6 wp 111 11 111     mt  other metal
A81a03 c1 wp 111 11 211     cv  canvas
A81a04 c1 wp 111 11 211     tx  textile
A81a05 b1 wp 111 11 111     pp  paper
A81a06 c2 wp 111 11 111     wp  wallpaper
A81a07 c2 wp 111 11 111     pg  pigment
A81a08 c2 wp 111 11 111     vn  varnish
A81a09 c2 wp 111 11 111     dy  dye
A81a10 c2 wp 111 11 111     ph  photograph
A81a11 c2 wp 111 11 111     dg  daguerreotype
A81a12 c2 lt 111 11 111     st  stone
A81a13 c6 wp 111 11 111     cm  ceramic
A81b01 a2 cu 444 33 411     gl  glass
A81b02 c1 pp 111 14 111     wd  wood
A81b03 b1 ph 444 31 412     lt  leather
A81b04 c3 ot 433 41 111     iv  ivory
A82a01 c3 tx 343 31 413     om  other organic
A82a02 c2 cu 111 11 111     at  analytical technique
A82a03 b3 dy 444 33 433     ot  other
A82a04 c3 pp 443 33 312
A82a05 c1 cv 111 11 111     Codes for Design, Data, Statistics
A82a06 a6 mt 111 41 111
A82b01 c3 pp 442 41 312     1   no; inapplicable
A82b02 c3 dg 442 13 311     2   no; should have
A82b03 c3 tx 323 13 431     3   yes; unclear or incorrect
A82b04 b4 lt 443 14 111     4   yes; clear and correct
A82b05 c1 at 111 11 111
A82b06 c1 at 111 11 111
A83a01 b6 pp 111 11 111
A83a02 c2 pg 111 11 111
A83a03 a3 pg 444 41 111
A83a04 c1 pp 111 11 111
A83a05 c1 ot 111 11 111
A83a06 c2 om 111 11 111
A83a07 c1 ot 111 11 111
```

A83b01 c6 ot 111 11 111
A83b02 c4 pp 442 33 313
A83b03 c6 ot 111 11 111
A83b04 a1 cu 111 44 411
A83b05 c4 pp 433 43 412
A83b06 b6 ag 111 11 111
A84a01 c6 pg 111 11 111
A84a02 a1 mt 111 41 111
A84a03 c1 pp 111 11 111
A84a04 b3 pg 444 14 441
A84a05 a2 pg 443 41 211
A84a06 c1 ot 111 11 111
A84b01 c6 cv 111 11 111
A84b02 c2 pg 111 11 111
A84b03 c2 pp 111 11 111
A84b04 a1 dy 424 41 111
A84b05 c4 dg 442 44 112
A84b06 c3 lt 443 11 111
A84b07 c1 cv 111 11 111
A85a01 c3 cv 443 34 412
A85a02 b3 pg 444 43 212
A85a03 c3 wd 444 41 444
A85a04 c6 pg 111 11 111
A85a05 a1 dy 442 31 111
B81a01 b3 cv 444 44 413
B81a02 a2 pg 443 44 411
B81a03 b2 st 111 11 111
B81a04 c4 fe 444 41 411
B81a05 b4 mi 423 41 111
B81b01 c4 dg 422 11 111
B81b02 b6 mi 111 11 111
B81b03 a2 pg 433 31 212
B81b04 c3 cu 422 41 111
B81b05 c1 cu 111 11 111
B81c01 a4 pg 443 41 412
B81c02 c6 ot 111 11 111
B81c03 b3 mt 444 41 112
B81c04 a1 tx 111 41 111
B81d01 c2 wd 444 41 111
B81d02 b1 mi 111 41 111
B81d03 b4 st 433 13 212
B81d04 c2 cm 111 11 111
B81d05 b6 pg 111 11 111
B81d06 b1 st 111 41 111
B82a01 a1 pg 111 44 111
B82a02 b2 st 444 44 411
B82a03 c3 pp 443 44 411
B82a04 a2 pg 443 41 111
B82a05 c3 st 442 24 342
B82a06 c4 fe 422 14 211
B82b01 a6 pg 111 11 111
B82b02 c1 ag 111 11 111
B82b03 b2 cu 111 11 111
B82b04 b3 vn 432 14 122
B82b05 c3 cm 433 44 111
B82b06 b4 mi 433 44 441
B82b07 a2 pg 111 11 111
B82c01 b4 fe 433 44 121
B82c02 b3 pg 433 14 112
B82c03 b3 iv 333 14 442
B82c04 a4 pg 443 44 111
B82c05 c3 wd 444 44 412
B82d01 c2 mt 422 44 411
B82d02 c2 tx 111 11 111
B82d03 c2 pp 111 11 111
B82d04 a2 wd 111 41 111
B82d05 c6 fe 111 11 111
B82d06 b2 cu 434 11 111
B83a01 c2 cu 441 34 312
B83a02 c3 tx 444 43 312
B83a03 a4 pg 444 41 111
B83a04 c6 at 111 11 111
B83a05 a4 mt 434 41 211
B83a06 b3 tx 433 34 223
B83b01 a1 pg 111 44 111
B83b02 b2 cu 444 13 131
B83b03 c3 st 432 34 412
B83b04 c1 at 444 44 412
B83b05 b4 fe 444 44 411
B83c01 c6 ot 111 11 111
B83c02 c4 pg 111 14 111
B83c03 a2 pg 444 44 111
B83c04 c1 pp 111 11 111
B83c05 a2 st 442 44 111
B83c06 c3 wd 332 33 442
B83d01 a4 dy 443 14 411
B83d02 c3 cm 443 44 112
B83d03 c3 st 423 11 111
B83d04 c1 at 342 44 441
B83d05 c1 wd 111 41 111
B83d06 b6 mt 111 11 111
B84a01 c2 tx 422 11 111
B84a02 a2 cm 442 44 111
B84a03 c1 cu 111 41 111
B84a04 c1 lt 111 11 111
B84a05 b2 st 433 11 111
B84a06 a1 pg 111 44 111
B84a07 c2 st 111 41 111
B84b01 b1 ph 111 11 111
B84b02 c3 gl 444 44 442
B84b03 c2 wd 111 14 111
B84b04 b1 dg 111 14 111
B84b05 b4 mt 443 34 111
B84b06 c2 om 111 11 111
B84c01 c1 at 443 44 111
B84c02 c3 pp 423 32 311
B84c03 c1 cv 111 11 111
B84c04 b1 st 111 41 111
B84c05 c1 cv 111 11 111
B84c06 a2 pg 444 44 111
B84c07 c1 pg 111 11 111
B84d01 c2 st 444 44 411
B84d02 b1 cu 111 14 111
B84d03 c2 wd 111 11 111
B84d04 a1 lt 111 41 111
B84d05 c1 st 422 13 431
B85a01 a1 pg 111 11 111
B85a02 a1 cv 111 11 111
B85a03 b6 fe 111 11 111
B85a04 c2 mt 432 41 111
B85a05 c2 cu 432 34 111
B85a06 a2 cm 434 34 111
B85a07 c2 wd 444 11 111
B85b01 b4 cu 444 44 131
B85b02 a1 dy 444 41 111
B85b03 c4 wd 443 44 112
B85b04 b4 st 443 44 111
B85b05 c6 lt 111 11 111
B85b06 a1 pg 111 44 111
B85b07 c6 ot 111 11 111
B85c01 c1 pg 443 11 111
B85c02 a2 cu 111 11 111
B85c03 a1 dy 433 44 111
B85c04 c2 pg 422 11 111
B85c05 c4 fe 443 44 431
B85c06 c4 pg 434 11 111
B85d01 a2 pg 444 41 111
B85d02 a4 pg 444 41 111
B85d03 b3 pg 444 44 412
B85d04 c1 cv 111 11 111
B85d05 b2 st 434 41 112
B85d06 b2 wd 444 41 411
B85d07 c2 om 444 11 111
C81a01 c1 at 111 44 111
C81a02 a6 pg 111 11 111
C81a03 c2 pg 111 11 111
C81a04 a2 pg 424 11 111
C81a05 a6 pg 111 11 111
C81a06 a2 pg 323 11 111
C81a07 a2 pg 411 11 111
C81a08 c2 wd 111 11 111
C81a09 a2 pg 424 11 111
C81a10 c3 cv 444 44 112
C82a01 a1 at 111 14 111
C82a02 a2 pg 444 11 111
C82a03 c2 wd 111 11 111
C82a04 a2 pg 424 11 111
C82a05 a2 pg 111 11 111
C82a06 a2 pg 434 41 111
C83a01 a2 pg 111 41 111
C83a02 a2 pg 443 41 111
C83a03 a2 pg 424 11 111

C83a04 a2 pg 424 11 111
C83a05 a6 pg 111 11 111
C83a06 c2 wd 111 11 111
C84a01 c1 at 111 44 111
C84a02 c2 pg 111 11 111
C84a03 a2 pg 422 11 111
C84a04 a2 wd 111 11 111
C84a05 c2 wd 111 11 111
C84a06 c2 ot 111 44 111
C84a07 c2 wd 111 11 111
C85a01 c1 at 111 44 111
C85a02 a2 pg 442 41 111
C85a03 a6 pg 111 11 111
C85a04 c2 pg 111 11 111
C85a05 a2 pg 444 11 111
C85a06 a6 pg 111 11 111
C85a07 c2 wd 111 11 111
C85a08 a4 pg 432 44 111
C85a09 b3 cv 444 44 112
D81a01 c6 mt 111 11 111
D81a02 c2 tx 111 11 111
D81a03 c2 pp 111 11 411
D81a04 c2 pg 111 11 111
D81a05 c3 tx 433 14 441
D81a06 b6 iv 111 11 111
D81a07 a4 pg 444 21 111
D81a08 a3 pg 434 14 131
D81a09 c1 ph 111 11 111
D81a10 c6 ot 111 11 111
D81a11 c6 ot 111 41 111
D81a12 c4 om 422 21 111
D81a13 b2 pg 111 11 111
D81a14 c2 cu 111 11 111
D81a15 a4 pg 222 41 111
D81a16 c6 pp 111 14 111
D81a17 a6 pg 111 11 111
D81a18 c2 wd 111 11 111
D81a19 c2 pg 422 11 111
D81a20 b4 dg 423 21 111
D81a21 a4 iv 423 11 111
D81a22 c2 mt 111 11 111
D81a23 c2 wd 111 11 111
D82a01 c6 pp 111 11 111
D82a02 c4 dg 333 41 411
D82a03 b2 cu 422 11 111
D82a04 b4 pp 443 44 412
D82a05 c4 pp 443 14 112
D82a06 c2 mt 111 11 111
D82a07 b4 pp 444 41 411
D82a08 c2 st 111 41 111
D82a09 c6 at 111 11 111
D82a10 c6 pg 111 11 111
D82a11 a4 pg 442 31 211
D82a12 c6 ot 111 11 111
D82a13 c1 pp 111 11 111
D82a14 c2 pp 111 11 111
D82a15 c2 cm 111 11 111
D82a16 a4 pg 444 41 211
D82a17 a1 vn 111 14 111
D82a18 c2 pg 111 11 111
D82a19 b4 ph 444 44 111
D82a20 c1 ot 111 14 111
D82a21 c2 wd 111 11 111
D82a22 a6 pg 111 11 111
D83a01 c1 ot 111 11 111
D83a02 a6 pg 111 11 111
D83a03 c6 ot 111 11 111
D83a04 c2 pp 111 11 111
D83a05 a4 pg 444 41 211
D83a06 c6 ot 111 11 111
D83a07 c4 mt 432 11 111
D83a08 b6 pg 422 44 131
D83a09 a2 wd 111 11 111
D83a10 a6 vn 111 11 111
D83a11 c2 cm 111 11 111
D83a12 c2 pg 111 11 111
D83a13 a6 ph 111 11 111
D83a14 c6 ot 111 11 111
D83a15 c2 pg 111 11 111
D83a16 c4 pp 443 41 412
D83a17 c2 pp 111 11 111
D84a01 c6 ot 111 11 111
D84a02 a6 pg 111 11 111
D84a03 a4 pg 444 41 221
D84a04 c6 pp 111 11 111
D84a05 a2 pg 434 44 111
D84a06 a1 ph 111 14 111
D84a07 c6 ot 111 11 111
D84a08 c2 pp 111 11 111
D84a09 c2 st 111 11 111
D84a10 c6 ot 111 11 111
D84a11 c2 pp 111 11 111
D84a12 c6 pp 111 14 111
D84a13 c6 ot 111 11 111
D85a01 b6 pp 111 11 111
D85a02 c2 pg 111 11 111
D85a03 c4 mt 433 31 111
D85a04 c2 pp 111 11 111
D85a05 c6 ot 111 11 111
D85a06 a2 pg 443 31 111
D85a07 c2 wd 111 11 111
D85a08 a1 pg 111 11 111
D85a09 c2 mt 111 11 111
D85a10 c6 gl 111 11 111
D85a11 c6 ot 111 11 111
D85a12 a6 om 111 11 111
D85a13 c1 pp 111 11 111
D86a01 c2 pg 111 11 111
D86a02 c2 tx 111 11 111
D86a03 c6 ot 111 11 111
D86a04 c6 ph 111 11 111
D86a05 c2 mt 111 11 111
D86a06 a2 pp 111 11 111
D86a07 a2 pg 444 41 111
D86a08 c2 pp 111 11 111
D86a09 b3 dy 444 34 111
D86a10 c2 ot 111 11 111

Figure 19.
Percentage of 320 papers in each category



each category. In this study, all variables were treated as being categorical or ordinal although year is clearly numerical. The issue and sequence identifiers were ignored for the rest of the analysis.

BMDP program 4F generates and analyzes frequency tables. In a frequency table, rows are labeled with the possible values of one variable and columns are labeled with the possible values of another variable. In this study, the cells in each row and column of the matrix contain the number of papers with both of the corresponding values of the row and column variables. The two variables are said to be cross-tabulated. Examples are found at the beginning of Chapters 2-4 and later in this chapter.

The classification variables were first cross-tabulated against each other. Then journal, year, and phase were tabulated against the eight statistical categories (number of research conditions, number of replicates, sampling design, tables, plots, descriptive statistics, estimation, and hypothesis testing). These were repeated using the subset of studies for which at least treatment number was appropriate. Both Pearson's Chi-square and Spearman's Rank Correlation tests were used as appropriate with $p < .05$ considered significant. (See Dixon 1985 for detailed descriptions of both BMDP programs.)

Survey Results and Discussion

Classification Variables

The data for the 320 papers reviewed is given in Figure 18. The percentage distribution of these papers among the categories of the classification variables journal, phase, type, and art material is presented in Figure 19. There is an even distribution across the years 1981 to 1985 with a couple percent in each of 1980 and 1986. The distribution of articles among the journals and years and the combinations thereof is a feature of our experimental design and does not need any further comment. Over half the papers in these journals focus on the conservation phase of a conservation project.

Coatings are by far the most common type of material category studied (106, or 33%). Pigments account for 83 of those papers. Of the remainder, dyes are studied in 7 papers, varnishes in 2, daguerreotypes in 7, and photographs in 6. The substrate category is the primary focus of 69 (22%) of the papers; these consist of 33 for paper, 12 for wallpaper, 13 for painting canvases, and 10 for textiles. Metals account for 44 papers (14%), with 16 papers about copper-based metals, 7 about archaeological iron, 4 about marine iron, 2 about silver, and 15 about other metals. Other organic materials make up 11% of the papers (37). The largest number of these are focused on wood (23), with 6 on leather, 3 on ivory, with the remaining 5 including the other organic materials. The mineral category consists of 26 papers (8%). Stone accounts for 15 of those papers with ceramics 7, glass 2, and shell 2. The "other" category described above accounts for the remaining 38 papers. Analytical techniques are the primary focus in 11 of those papers, and general conservation and management issues account for the others.

Interaction of Classification Variables

As noted above, the exact number of each journal and year is set by design. Because we reviewed only one issue of one journal in each of 1980 and 1986, there is an "interaction" between journal and year which is, however, purely a characteristic of our design rather than of the conservation literature.

Phase, type, and material are somewhat different variables since their values are measured. The only similarity to journal and year is that we made some effort to choose categories that would result in an approximately even distribution.

There is a major difference in analysis of categorical as opposed to quantitative variables. Analysis of variance of a quantitative measure looks at the mean values of items that have the same combination of applied treatments. We are interested in first order effects and usually prefer that there be no interaction effects.

As an example, we might test two pigments in two types of media (oil and acrylic). We could then measure the degree of color change (yellowing and fading) that occurs after accelerated aging of several replications of each of the four pigment-medium combinations and calculate the four means. We could then ask and get answers to three questions:

M. Does one medium yellow significantly less than the other?
P. Does one pigment type fade significantly less than the other?
MxP. Does the medium effect depend upon the pigment, and vice-versa?

Questions M and P are about first-order effects. Question MxP is about a second-order or interaction effect. We would prefer that the interaction effect be negligible so that we could conclude that one medium is significantly better than another regardless of the pigment type.

In a contingency table analysis, we analyze the number of objects falling in the cells of the matrix defined by the possible values of two or more measured categories. The categories listed on the side of the table, rather than being types of treatments, are the categories into which the articles fall. The presence of first-order effect means that for a given variable the number of objects falling in the different categories of that variable is uneven. This is generally of little interest unless there is some prior expectation of an even distribution. We are usually more concerned with interactions between variables.

Phase, type, and material are measured variables and are a product of our design only insofar as we have chosen the categories. The imbalance between different phases is of some interest, although we could make it look more even by combining composition and deterioration. Counts of article types are not important, as they were basically chosen to be even, particularly when the environmental studies were combined with other general studies. Material types are also fairly even, especially after the categories were collapsed to give enough counts to each group to allow for meaningful statistical analysis.

The interesting aspect of these variables are their interrelationships and relation to the statistical variables. There are some significant relationships between journal and article category. Figure 20 shows that

TB contains more art composition studies than conservation studies, which is to be expected since its focus is technical studies of art objects rather than conservation per se. All of the other journals contain more conservation studies than anything else, with JC and PP being about two-thirds studies of conservation materials and methods. SC has the most even distribution.

Figure 20.
Percentage of papers in each journal concerning each phase

Journal	Composition	Deterioration	Conservation
SC	24	28	48
JC	16	12	72
PP	24	10	66
TB	60	3	37

There are also some significant relationships between journal and article type. Figure 21 shows the percentage in each journal of each article type. TB is heavily weighted towards case studies, and PP to essays. SC has the most even spread. JC is also fairly even, but has somewhat more essays and papers concerned with how to carry out a particular procedure.

Figure 21.
Percentage of papers in each journal of each type

Journal	How-to	Case Study	Gen(Sim)	Gen(Real)	Essay
SC	26	31	16	18	9
JC	32	22	20	8	18
PP	8	39	3	17	33
TB	11	68	5	3	13

There also is a significant interaction between phase and type. The distribution of types is given at the beginning of the chapters on each phase. There do not seem to be any significant relationships between year and phase, type, or material. In other words, there are no temporal trends in these latter three variables.

Statistical Variables

In 184 out of 320 papers it did not make sense to talk about the number of treatments. If treatment number was not applicable, then no other statistical categories were either. Thus these 184 were coded 1 (inapplicable) for all the statistical categories and are not considered further. We will restrict our attention to the subset of 136 papers for which treatment number was applicable. The distribution of assigned categories for each variable for the papers in this subset is given in Figure 22.

Figure 22.
Number of papers in each statistical category among papers for which treatment number was applicable

Statistical Category	Inapplicable or Not Done	Should Have Done	Unclear	Clear
Treatment Number	0	1	7	128
Replicate Number	1	27	30	78
Sampling Design	2	31	47	56
Table	17	4	22	93
Plot	47	1	14	74
Descriptive Statistics	76	13	8	39
Estimation	115	4	8	9
Hypothesis Testing	99	31	6	0

This table highlights the major findings of our survey. The number of papers clearly reporting successive statistical aspects goes down at increasing levels of statistical sophistication.

Most papers in art conservation research are not designed to produce quantitative, testable data. Thus a minority of 42% fall in our subset with statistics potentially applicable, and only for a minority of these were any of the three analysis variables actually applicable. In many of these cases, however, this evaluation could be changed by redesigning the study.

Hypothesis testing should be a standard statistical procedure in art conservation research, particularly in the general studies that account for 26% (84) of the journal papers published over the past five years. However, only 37 papers were designed to collect data amenable to hypothesis testing. In 31 of those papers, no attempt at hypothesis testing was reported; interpretations and conclusions were apparently based on a qualitative, visual perusal of the data. In none of the remaining six were both execution and presentation of hypothesis testing completely satisfactory. In some cases we simply could not tell how the analysis was done (i.e., what was taken to be a repeated measures factor and what was taken to be a grouping factor). In other cases, clearly inappropriate methods were used (such as doing multiple t-tests or ignoring correlation and treating a repeated measures factor as if it were a grouping factor, both of which tend to give spuriously high significance levels).

Correlation of Classification and Statistical Variables

Studies in Conservation has the highest degree of statistical applicability because its case studies tend to include some experimental work. The clear reporting of replicate and repeated measure number improved from 37% in 1981 to 69% in 1985. There are no other significant relationships between journal or year and the statistical variables. The only relationship for phase is that 33% of the deterioration and conservation articles in the subgroup of 136 should have done a statistical test while the corresponding figure for composition articles is 4%.

Appendix

The title of each section of this appendix gives the section of the text with the corresponding discussion. Each contains the data, BMDP input, and BMDP output for the analysis of a particular study. The actual computer files are printed in the smaller fixed-spacing type. Any text following a "#" on an input file line is ignored by BMDP as a comment. The output has been condensed and edited from its original version and an occasional comment added using the same convention for "#."

The BMDP Statistical Software package consists of several programs which are identified by two letter codes. We have used the following programs from this package:

- 1D Means, standard deviations, minimum and maximum for each variable
- 5D Histograms for individual variables
- 6D Scattergram plots with options for groups and multiple variables
- 7D One or twoway analysis of variance with small histograms for each group
- 8D Correlation matrices
- 2V Analysis of variance with repeated measures and covariates

When starting a new project, our general practice is to use 1D first to get a one-line summary for each variable and generate an analysis file (with /SAVE) in BMDP's internal format for use by the other programs. The condensed summary of what the program read and how it interpreted it is useful for checking that the data have been entered correctly and that they have been properly described to the program. The only exceptions are when the data matrix is small, has a simple structure, and will be needed for only one analysis. Otherwise, our experience shows that it is faster to proceed in manageable steps rather than trying to do everything at once in one computer run.

Statistical programs require that the input data be properly organized. Nearly always, the rows or lines must represent the object or entity being analyzed, while the data columns or variables represent the attributes and properties of those objects. The major exception for BMDP is that nonlinear regression analysis of repeated measures with programs 3R or AR requires a transposed matrix with columns representing objects and rows representing the repeated measures. The important point is to think about what analyses will be needed or consult with a data analyst or statistician before doing experiments so the data can be recorded in the most useful format.

Statistical programs also have to be able to separate the characters on a line into distinct values, one for each variable. This can be done either by assigning one or more single-character columns to each variable or by separating the values with a space, comma, or other special marker. The files in this appendix use both methods. There are few enough variables to use spaces as separators and still keep everything on one line,

which makes the format easy to describe ("FORMAT = FREE." in the BMDP input). Keeping values in vertical alignment, although a bit of extra effort, makes them easier to read and check.

Lastly, if any values are missing, their absence must be indicated by an appropriate place holder. The files in this appendix are all complete so we can ignore this problem here.

A.1

Pigment Palette (England and van Zelst 1982)

```
# 8D
# Palette study of 17th New England portrait paintings. (pp. 92,94)
# The variable names are abbreviations of those used in Table 2-3.
#################################################################################

# Rows represent paintings, columns represent pigments

/INPUT  var   = 11.
        form  = free.
/VAR    name  = yellake, redlake, ltyellow, vermilon, curesin,
                grnearth, ultramar, realgar, smalt, umber, gold.
/PRINT  level = brief.
/END
0 0 0 0 0 0 0 0 1 0 1
0 1 1 1 1 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 1
0 1 1 1 0 0 0 0 0 0 0
0 1 1 1 0 1 0 0 0 1 0
0 1 1 1 1 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 0 1 1 0 0
0 0 0 1 1 1 0 0 1 0 0
0 0 1 1 1 0 0 0 0 0 0
0 0 1 1 1 1 0 0 1 0 0
0 1 1 1 1 0 0 0 1 1 0
0 1 0 1 0 1 1 0 0 0 0
0 1 1 0 0 0 0 0 1 0 0
1 1 0 1 0 0 0 0 0 0 0
end

# Reverse row and columns so that row=pigment, column=painting

/INPUT  var   = 15.
        form  = free.
/VAR    name  = bonner, smith, gibbs, mason, pattesh, eggingtn,
                freake, winthrop, downing, savage, jwensley,
                ewensley, dark, davenprt, rawson.
/PRINT  level = brief.
/END
0 0 0 0 0 0 1 1 0 0 0 0 0 1 0
1 1 1 1 1 1 1 1 0 1 1 1 0 0 0
1 1 1 1 0 1 1 1 1 0 0 0 0 0 0
1 0 0 0 0 1 1 0 0 1 1 1 0 0 0
0 0 0 1 1 1 0 1 0 1 1 1 0 0 0
0 0 1 1 1 1 1 1 1 1 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 1 1 1 1 1 1 0 0
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0 0 1 1 0
```

A.2

Lead Isotopes (Brill, Barnes,andMurphy1981)

The following ratio data is taken from their Table 1. The sample id is followed by the significant digits of the Pb 208, 207, and 204 ratios to 206.

```
 616  676  339  312        622  784  389  350
 617  687  341  318       1321  791  392  366
 618  693  341  311        701  770  388  346
 416  694  352  327        646  763  390  354
  46  721  340  316        299  770  393  355
 650  699  346  321        228  745  392  356
 651  710  348  322        620  818  341  318
1202  714  348  321         40  819  354  376
1023  681  352  326        623  814  370  341
  85  723  352  322        215  828  374  342
1330  721  355  339        285  832  375  346
 733  717  354  330       1011  825  382  330
 673  719  357  309        658  843  386  372
 729  716  359  335        724  812  388  353
1252  715  360  335        415  810  400  362
 721  746  354  330       1010  825  407  360
 730  760  364  341        204  812  408  361
1319  736  366  328        431  807  412  365
1316  710  372  332        722  818  414  373
1387  723  374  344        636  809  417  374
1201  731  375  340        289  834  422  385
 652  744  374  335        418  822  424  386
 664  754  377  334        703  819  424  411
1320  762  376  344        662  817  425  378
 430  764  380  354        626  832  432  390
 222  759  382  351        637  840  435  382
1315  784  372  330        732  778  424  386
 621  778  386  347        601  941  497  421
```

```
# 1D Read significant figures of ratios and convert to fractions.
# Given ratios a/d, b/d, c/d and relation a + b + c + d = 1,
# then d = 1 / (a/d + b/d + c/d + 1) and a = (a/d) * d, etc.
#########################################################################################
/INPUT  var   = 4.
        form  = '(a4, 3i5)'.
        file  = 'ratio.data'.
/VAR    name  = id, pb8_6, pb7_6, pb4_6.
        add   = new.
        label = id.
/TRANS  pb8_6 = 2.0  + .0001  * pb8_6.
        pb7_6 = 0.8  + .0001  * pb7_6.
        pb4_6 = 0.05 + .00001 * pb4_6.
        pb206 = 1 / (pb8_6 + pb7_6 + pb4_6 +1 ).
        pb204 = pb4_6 * pb206.
        pb207 = pb7_6 * pb206.
        pb208 = pb8_6 * pb206.
/PRINT  data.       level = brief.
/SAVE   new.        file  = save.        code  = biomath.
/END
```

While "/PRINT data." in the 1D program above causes the fractions to be printed, there is no control over the format. The following C program gives a more readable listing. A Fortran or Basic program would be similar.

```
#include <stdio.h>

main () {
  double pb8_6, pb7_6, pb4_6, pb204, pb206, pb207, pb208;
  while (scanf("%*4c%5lf%5lf%5lf\n", &pb8_6, &pb7_6, &pb4_6) == 3) {
        pb8_6 = 2.0  + .0001  * pb8_6;
        pb7_6 = 0.8  + .0001  * pb7_6;
        pb4_6 = 0.05 + .00001 * pb4_6;
        pb206 = 1.0 / (pb8_6 + pb7_6 + pb4_6 + 1);
        pb204 = pb4_6 * pb206;
        pb207 = pb7_6 * pb206;
        pb208 = pb8_6 * pb206;
        printf("%7.6f %7.6f %7.6f %7.6f %5.1f %5.1f\n", pb204, pb206, pb207, pb208;
   }
}

# 5D Histograms
################################################################################
/INPUT   file  = save.        code  = biomath.
/PLOT    var   = pb8_6 to pb208.
/PRINT   level = brief.
/END

# 8D Correlation
################################################################################
/INPUT   file  = save.        code  = biomath.
/CORR    row   = pb8_6 to pb208.
         col   = pb8_6 to pb208.
/PRINT   level = brief.
/END

# 6D Plot combinations of ratio and fractions and ternary plot.
################################################################################
/INPUT   file  = save.        code  = biomath.
/VAR     add   = new.         group = pb204.
/TRAN    use   = kase ne 56.                    _
         terx  = (1 - pb206 + pb207) * . # = 1 / V3
/GROUP   cutp(pb204) = .
/PLOT    xvar  = pb7_6, pb4_6, pb4_6, pb207, pb206, pb204, pb206, pb204, pb204.
         yvar  = pb8_6, pb8_6, pb7_6, pb208, pb208, pb208, pb207, pb207, pb206.
         pair.
         size  = 90, 48.
/PLOT    xvar  = terx.
         yvar  = pb208.
         symbol= '.', '-', '+', '*', '#'.
/PRINT   level = brief.
/END
```

A.3

Densitometer (Wilhelm 1981)

```
# 2V Test densitometer and filter effect with film as subject.
#################################################################################################

/INPUT   var   = 9.
         form  = '(9i3)'.
/VAR     name  = red1, green1, blue1,
                 red2, green2, blue2,
                 red3, green3, blue3.
/DESIGN depend= red1 to blue3.
         level = 3, 3.
         name  = dens, col.
/PRINT   level = brief.
/END
 72 84 74 68 81 74 67 75 68
 82 84 78 78 82 80 77 74 73
 66 76 64 59 72 63 61 67 60
 85 97101 75 88 99 69 80 96
 77 94 80 72 91 75 98 86 70
end
```

CELL	dens	col	MEAN
red1	1	1	76.4
green1	1	2	87.0
blue1	1	3	79.4
red2	2	1	70.4
green2	2	2	82.8
blue2	2	3	78.2
red3	3	1	74.4
green3	3	2	76.4
blue3	3	3	73.4

```
# standard deviations range from 7.1 to 14.4

ANALYSIS OF VARIANCE FOR Red1 green1 blue1 red2 green2 blue2 red3 green3 blue3
```

SOURCE	SUM Of SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
MEAN	70979.20000	1	270979.20000	407.94	.000
ERROR	2657.02222	4	664.25556		
dens	293.20000	2	146.60000	5.99	.026
ERROR	195.91111	8	24.48889		
col	528.93333	2	264.46667	2.20	.173
ERROR	962.17778	8	120.27222		
dc	185.86667	4	46.46667	2.27	.107
ERROR	327.68889	16	20.48056		

The line labeled 'MEAN' says, in this case, that the mean density is significantly different from 0. When this is assumed, as is usual, the MEAN line can be ignored. Exceptions occur if, for instance, a difference is being analyzed.

The line labeled 'dens' says that the probability of the observed variation attributable to the densitometer effect, under the null hypothesis of no densitometer differences, is .026, which is usually considered statistically significant since less than .05, although this number (.05) is not quite as magical as sometimes made out to be.

The null hypothesis probabilities for the effect of different color filters and the interaction between densitometer and filter are much larger and would usually be interpreted to indicate that these effects are not likely to be significant.

```
# 2V  Test film and filter effect with densitometer as subject.
# Note that data must be rearranged for the second analysis.
##############################################################################################

/INPUT  var   = 15.
        form  = '(15i3)'.
/VAR    name  = red1, green1, blue1, red2, green2, blue2,
                red3, green3, blue3, red4, green4, blue4,
                red5, green5, blue5.
/DESIGN depend= red1 to blue5.
        level = 5, 3.
        name  = film, col.
/PRINT  level = brief.
/END
 72 84 74 82 84 78 66 76 64 85 97101 77 94 80
 68 81 74 78 82 80 59 72 63 75 88 99 72 91 75
 67 75 68 77 74 73 61 67 60 69 80 96 98 86 70
end

# output for second problem

CELL                    MEAN
            film col
 red1        1    1     69.0     # The number of objects in each cell is 3.
 green1      1    2     80.0     # The mean for all cells if 77.6.
 blue1       1    3     72.0     # Standard deviations range from 2.1 to 13.8.
 red2        2    1     79.0
 green2      2    2     80.0
 blue2       2    3     77.0
 red3        3    1     62.0
 green3      3    2     71.7
 blue3       3    3     62.3
 red4        4    1     76.3
 green4      4    2     88.3
 blue4       4    3     98.7
 red5        5    1     82.3
 green5      5    2     90.3
 blue5       5    3     75.0

SOURCE        SUM OF     DEGREES OF       MEAN           F       TAIL
              SQUARES     FREEDOM         SQUARE                 PROB.

MEAN       70979.20000       1       270979.20000     1848.43     .001
ERROR        293.20000       2          146.60000

film        2657.02222       4          664.25556       27.12     .000
ERROR        195.91111       8           24.48889

col          528.93333       2          264.46667        5.69     .068
ERROR        185.86667       4           46.46667

fc           962.17778       8          120.27222        5.87     .001
ERROR        327.68889      16           20.48056
```

A.4

Pigments (Simunkova et al. 1985)

```
# 2V Analysis of covariance of days by pigment with concentration covariate.
#################################################################################
/INPUT  var   = 3.
        form  = free.
/VAR    name  = pigment, conc, days.
/GROUP  code(pigment) = 1,2,3,4.
/DESIGN depend= days.
        group = pigment.
        cova  = conc.
/PRINT  level = brief.
/END
1  5 19.5
1 10 18
1 20 15
1 30 10
2  5 14
2 10 13
2 20  7
2 30  5
3  5 15
3 10 16
3 20  9
3 30  6.5
4  5  9
4 10  6
4 20  3
4 30  2.5

CELL MEANS FOR  1-ST COVARIATE                                  MARGINAL
pigment =  *1.00000    *2.00000    *3.00000    *4.00000
conc        16.25000    16.25000    16.25000    16.25000    16.25000

STANDARD DEVIATIONS
conc        11.08678    11.08678    11.08678    11.08678

CELL MEANS FOR  1-ST DEPENDENT VARIABLE                         MARGINAL
pigment =  *1.00000    *2.00000    *3.00000    *4.00000
days        15.62500     9.75000    11.62500     5.12500    10.53125

STANDARD DEVIATIONS
days         4.19076     4.42531     4.60751     3.01040
COUNT           4           4           4           4          16

ANALYSIS OF VARIANCE FOR DEPENDENT VARIABLE - days

SOURCE       SUM OF    DEGREES OF      MEAN          F      TAIL    REGR
             SQUARES    FREEDOM       SQUARE               PROB.  COEFF
pigment     227.92188      3         75.97396       45.05    .000
conc        183.76298      1        183.76298      108.97    .000  -.353
ERROR        18.54952     11          1.68632
```

A.5

Fading and Dye Mordants (Crews 1982)

```
# 1D
# Color measure (change for E) at end of 4th exposure period.
# Each value is the mean of 2 replicate samples (pp 54-56).
###############################################################################
/INPUT   var   = 3.
         form  = free.
         file  = crews.data.
/VAR     name  = deltaE, litefast, grayscal.
         add   = new.
/TRAN    mordant = kase mod 5.
         dye     = kase mod 17.
/GROUP   name(mordant) = tin, alum, chrome, iron, copper.
         code(mordant) = 0, 1, 2, 4, 3.
         name(dye) = SourCher, ChokCher, Clover, Coreop, CrabAppl, Dock,
                     Fustic, Goldrod, Grape, Marigold, Mimosa, Mullein,
                     Onion, Peach, Poplar, Smartwed, Tumeric.
         code(dye) = 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,0.
/PRINT   level = brief.
/SAVE    new.
         file  =  'crews.save'.
         code  = biomath.
/END
 7.3 2 1.5         1.9 4 2.0         1.8 5 3.5        14.3 2 1.5
 2.9 3 2.0         2.9 4 2.0        18.1 2 1.0        11.8 4 2.0
 1.2 5 3.5        15.7 3 1.0        18.9 2 1.5         2.5 4 3.0
 2.7 4 3.0        11.8 3 1.5         4.8 4 2.5         1.0 5 3.0
 9.0 3 2.5         1.5 5 3.5         0.9 5 3.5         2.1 4 2.5
 5.3 3 2.0         1.0 5 3.0         4.7 3 1.5        10.3 4 2.5
 3.7 4 2.0         2.0 4 2.5        13.0 3 1.0        14.4 2 1.5
 0.9 5 3.5        16.4 2 1.5        17.5 2 1.5         2.1 4 2.5
 3.7 4 2.0        16.7 3 2.0         2.1 6 4.0         1.7 5 3.0
10.4 4 2.5         6.3 6 4.0         2.5 5 3.0         2.7 4 1.5
 8.5 4 2.0         0.3 7 5.0         2.3 4 2.0        15.8 3 1.5
 2.4 4 3.0         0.7 7 5.0        12.4 5 3.0        11.4 2 2.0
 0.8 6 4.0        20.9 2 2.0         9.7 3 2.0         1.7 4 2.5
 1.4 5 3.0        12.5 3 2.0         2.2 5 3.5         1.5 6 4.0
13.1 3 1.5         3.3 5 3.0         1.9 5 4.0         3.3 4 2.5
 7.6 2 1.5         2.3 5 3.5         1.7 5 3.5        12.3 2 1.5
 6.3 3 1.5         2.1 5 2.5         9.2 4 2.5        10.6 3 1.0
 3.3 4 2.0        16.4 3 1.5        13.9 3 1.5         6.4 3 1.5
 2.4 4 2.5         8.7 2 1.5         3.5 4 1.5         4.0 4 1.5
 9.3 2 1.5         2.1 6 4.0         2.2 5 3.5         2.1 5 3.0
15.5 2 1.5         0.6 7 4.5         1.7 5 3.0        23.5 2 1.0
 2.5 4 2.0        # There is only one set of columns in original file.

VARIABLE                          STANDARD    ST.ERR    SMALLEST   LARGEST
NAME        FREQUENCY      MEAN   DEVIATION   OF MEAN     VALUE     VALUE

deltaE         85          6.668     5.906      .6406       .300    23.500
litefast       85          3.906     1.306      .1416      2.000     7.000
grayscal       85          2.424      .968      .1050      1.000     5.000
```

```
# 8D Correlation (Results are included in the text discussion.)
#################################################################################################

/INPUT   file  = 'crews.save'.
         code  = biomath.
/CORR    row   = deltaE, litefast, grayscal.
         col   = deltaE, litefast, grayscal.
/PRINT   level = brief.
/END


# 2V Analysis of variance
#################################################################################################

/INPUT   file  = 'crews.save'.
         code  = biomath.
/DESIGN depend= deltaE, litefast, grayscal.
         level = 1.
         group = mordant, dye.
         exclud= 12.
/PRINT   level = brief.
/END


ANALYSIS OF VARIANCE FOR  1-ST  DEPENDENT VARIABLE - deltaE

SOURCE          SUM OF     DEGREES OF       MEAN          F       TAIL
                SQUARES     FREEDOM        SQUARE                PROB.

MEAN          3779.55576       1        3779.55576      514.43     .0000
mordant       2342.22069       4         585.55517       79.70     .0000
dye            117.93224      16           7.37077        1.00     .4651
ERROR          470.21128      64           7.34705

ANALYSIS OF VARIANCE FOR  2-ND DEPENDENT VARIABLE - litefast

MEAN          1296.75294       1        1296.75294     1531.55     .0000
mordant         81.01176       4          20.25294       23.92     .0000
dye              8.04706      16            .50294         .59     .8769
ERROR           54.18824      64            .84669

ANALYSIS OF VARIANCE FOR  3-RD  DEPENDENT VARIABLE - grayscal

MEAN           499.24706       1         499.24706      812.05     .0000
mordant         33.75294       4           8.43824       13.73     .0000
dye              5.65294      16            .35331         .57     .8914
ERROR           39.34706      64            .61480
```

```
# 7D - Histograms of groups with anova
##################################################################################################

/INPUT  file = 'crews.save'.    code  = biomath.
/HIST   var  = deltaE, litefast, grayscal.
        group = mordant.
/PRINT  level = brief.   /END

HISTOGRAM OF * deltaE   * (   1)   GROUPED  BY   * mordant  * (   4)

         tin          alum         chrome       iron         copper
MIDPOINTS.....+.....+.....+.....+.....+.....+.....+.....+.....+.....
  27.000)
  25.500)
  24.000)*
  22.500)
  21.000)*
  19.500)            *
  18.000)*           *
  16.500)***         *
  15.000)**          **
  13.500)M*          *
  12.000)**          M***
  10.500)**          *
   9.000)***         ***
   7.500)            **
   6.000)            *            ***
   4.500)                         *            *            *
   3.000)                         M******      M******      ***
   1.500)                         ******       ********     M**********
    .000)                                       *            **
GROUP MEANS ARE DENOTED BY M'S

MEAN      14.124       11.888       3.312        2.371        1.647
STD.DEV.   4.155        3.863       1.651        0.931        0.989
S. E. M.   1.008        0.937       0.401        0.226        0.240
MAXIMUM   23.500       18.900       6.400        4.700        4.000
MINIMUM    9.000        5.300       1.500        0.700        0.300
CASES       17           17          17           17           17

-------------------------------------------------   ALL GROUPS COMBINED
|     ANALYSIS OF VARIANCE TABLE FOR MEANS     | EXCEPT CASES WITH UNUSED
|SOURCE  SUM OF SQUARES DF MEAN SQUARE  F PROB|   VALUES FOR mordant
|------- ------------- -- ----------- -- ----|  MEAN               6.668
|mordant    2342.2207    4    585.5552 80 .000|  STD. DEV.          5.906
|  ERROR     588.1435   80      7.3518       |  S. E. M.           0.641
|-----------------------------------------------|  MAXIMUM           23.500
| MEANS, VARIANCES ARE NOT ASSUMED TO BE EQUAL |  MINIMUM            0.300
|   WELCH               4,38    59.00     .000|  CASES EXCLUDED    (   0)
|   BROWN-FORSYTHE      4,41    79.65     .000|  CASES INCLUDED        85
|-----------------------------------------------|
| VARIANCES, LEVENE     4,80    11.31     .000|  ROBUST S.D.        6.433
-------------------------------------------------
```

A.6

Fading and Light Filters
(Bowman and Reagan 1983)

```
# 1D
# 3 dyes, 3 lamps, filtered or not, and 4 exposures times.
# Data are read from their plots (pp. 41,42)
# except that indigo values are 1 less than value on plot.
# K/S values are transformed to differences from initial value
#################################################################################

/INPUT  var   = 7.
        form  = '(3i1, 4f3.2)'.
/VAR    name  = dye, light, filter, h100, h200, h300, h400.
        group = dye.
/TRANS  FOR d = 1, 2, 3.
         x = 1.12, .80, 1.09. %
         if (dye eq d) then (
          FOR hour = 50, 100, 200, 400. % h|hour = x - h|hour. %
         ) .
        %
/GROUP  name(dye)   = tumeric, madder, indigo.
        code(dye)   = 1, 2, 3.
        name(light) = floures, quartz, incandes.
        code(light) = 1, 2, 3.
        name(filter)= bare, filtered.
        code(filter)= 0, 1.
/PRINT  level = brief.
        line  = 80.
/SAVE   new.
        file  = 'bowman.save'.
        code  = biomath.
/END
110 82 71 64 57
111 87 75 67 65
120 80 77 65 57
121 93 83 73 65
130 91 82 65 61
131 93 80 73 65
210 78 72 71 71
211 79 79 79 79
220 77 76 72 72
221 79 80 76 74
230 75 73 68 66
231 80 77 72 67
310 68 53 37 37
311 72 55 49 48
320 90 80 28 31
321 98 85 27 34
330 81 60 30 19
331 77 56 43 30
```

The summaries are given for each dye separately as well as all dyes combined. This was requested because the scale of differences was clearly smaller for madder than the other two dyes. A check of smallest/largest values can reveal gross entry errors. A check of the frequency tables to make sure that the right number is listed for each group is also vital.

VARIABLE NO.	NAME	GROUPING VAR/LEVL	TOTAL FREQ.	MEAN	STANDARD DEVIATION	SMALLEST VALUE	SMALLEST Z-SC	LARGEST VALUE	LARGEST Z-SC
4	h100		18	.181	.137	.000	-1.32	.410	1.67
		dye							
		tumeric	6	.243	.056	.190	-.94	.320	1.36
		madder	6	.020	.018	.000	-1.12	.050	1.68
		indigo	6	.280	.113	.110	-1.51	.410	1.15
5	h200		18	.273	.194	.000	-1.41	.560	1.48
		dye							
		tumeric	6	.340	.046	.290	-1.10	.410	1.53
		madder	6	.038	.032	.000	-1.20	.080	1.31
		indigo	6	.442	.140	.240	-1.44	.560	.85
6	h300		18	.415	.285	.010	-1.42	.820	1.42
		dye							
		tumeric	6	.442	.041	.390	-1.25	.480	.93
		madder	6	.070	.039	.010	-1.54	.120	1.28
		indigo	6	.733	.089	.600	-1.49	.820	.97
7	h400		18	.449	.292	.010	-1.50	.900	1.54
		dye							
		tumeric	6	.503	.039	.470	-.85	.550	1.19
		madder	6	.085	.048	.010	-1.57	.140	1.15
		indigo	6	.758	.095	.610	•1.56	.900	1.49

VARIABLE NO.	NAME	CATEGORY NAME	CATEGORY FREQUENCY	TOTAL FREQUENCY	NO. OF VALUES MISSING OR OUTSIDE THE RANGE
1	dye			18	0
		tumeric	6		
		madder	6		
		indigo	6		
2	light			18	0
		floures	6		
		quartz	6		
		incandes	6		
3	filter			18	0
		bare	9		
		filtered	9		

```
# 2V  Analysis of repeated measures by dye, light, and filter.
# In the output, h(1), h(2), and h(3) refer to separate linear,
# quadratic, and cubic time trends, as requested by 'orthogonal'.
#################################################################################
/INPUT    file  = 'bowman.save'.
          code  = biomath.
/DESIGN   depend= h50 to h400.
          level = 4.
          name  = hour.
          orthogonal.
          group = dye, light, filter.
          exclud= 123.
/PRINT    level = brief.
/END
```

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
MEAN	7.82101	1	7.82101	8004.45	.0000
dye	3.09923	2	1.54961	1585.96	.0000
light	.01916	2	.00958	9.80	.0287
filter	.04351	1	.04351	44.53	.0026
dl	.03809	4	.00952	9.75	.0244
df	.00061	2	.00030	.31	.7488
lf	.00206	2	.00103	1.05	.4291
ERROR	.00391	4	.00098		
h(1)	.80372	1	.80372	526.02	.0000
h{1)d	.33931	2	.16965	111.04	.0003
h(1)l	.05196	2	.02598	17.01	.0111
h(1)f	.00240	1	.00240	1.57	.2781
h(1)dl	.06296	4	.01574	10.30	.0221
h(1)df	.00210	2	.00105	.69	.5537
h(1)lf	.00529	2	.00265	1.73	.2872
ERROR	.00611	4	.00153		
h(2)	.01531	1	.01531	108.62	.0005
h(2)d	.01456	2	.00728	51.64	.0014
h(2)l	.00161	2	.00080	5.70	.0674
h(2)f	.00007	1	.00007	.48	.5254
h(2)dl	.00256	4	.00064	4.54	.0861
h(2)df	.00139	2	.00069	4.92	.0836
h(2)lf	.00042	2	.00021	1.49	.3288
ERROR	.00056	4	.00014		
h(3)	.02225	1	.02225	35.84	.0039
h(3)d	.02583	2	.01292	20.81	.0077
h(3)l	.02566	2	.01283	20.67	.0078
h(3)f	.00173	1	.00173	2.79	.1700
h(3)dl	.03644	4	.00911	14.68	.0117
h(3)df	.00094	2	.00047	.75	.5270
h(3)lf	.00226	2	.00113	1.82	.2736
ERROR	.00248	4	.00062		

A.7

Linen Canvas Strength
(Hackney and Hedley 1981)

```
# 1D
##############################################################################################
/INPUT   var   = 6.
         form  = '(4i1, f5, f4)'.
/VAR     name  = board, wax, dark, closure, strength, ph.
/GROUP   name(board) = '1', '2', '3'.
         code(board) = 1,2,3.
         name(wax)    = bare,   waxed.
         name(dark)   = light,  dark.
         name(closure)= open, closed.
         code(wax, dark, closure) = 1,2.
/PRINT   level = brief.
/SAVE    new.
         file  = 'hackney.save'.
         code  = biomath.
/END
1111  1.21  4.0
1112  2.27  5.3
1121  1.99  4.1
1122  2.61  5.5
1211  2.08  4.8
1212  1.90  4.8
1221  2.48  4.9
1222  2.50  5.1
2211  2.21  4.9
2212  1.93  4.3
2221  2.60  5.1
2222  2.21  5.2
3111  1.12  4.1
3112  2.12  5.7
3121  1.64  4.5
3122  2.50  5.9
3211  2.35  4.8
3212  2.09  5.1
3221  2.43  5.2
3222  2.25  5.4
```

VARIABLE NO.	NAME	TOTAL FREQUENCY	MEAN	STANDARD DEVIATION	ST.ERR OF MEAN	SMALL VALUE	LARGE VALUE
5	strength	20	2.125	.414	.0926	1.120	2.610
6	ph	20	4.960	.509	.1139	4.000	5.900

board	1	8
	2	4
	3	8
wax	bare	8
	waxed	12
dark	light	10
	dark	10
closure	closed	10
	open	10

```
# 2V    ANALYSIS OF VARIANCE AND COVARIANCE WITH REPEATED MEASURES.
# Exclude board in order to have an error term.
# If board in included, something else must be excluded.
##################################################################################################

/INPUT   file  = 'hackney.save'.
         code  = biomath.
/DESIGN depend= strength, ph.
         level = 1.
         group = wax, dark, closure.
/PRINT   level = brief.
/END

# Cell means are in text
```

ANALYSIS OF VARIANCE FOR 1-ST DEPENDENT VARIABLE - strength

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
MEAN	84.06828	1	84.06828	4930.69	.0000
wax	.49152	1	.49152	28.83	.0002
dark	.81345	1	.81345	47.71	.0000
closure	.54405	1	.54405	31.91	.0001
wd	.04181	1	.04181	2.45	.1433
wc	1.44321	1	1.44321	84.65	.0000
dc	.01633	1	.01633	.96	.3470
wdc	.03605	1	.03605	2.11	.1716
ERROR	.20460	12	.01705		

ANALYSIS OF VARIANCE FOR 2-ND DEPENDENT VARIABLE - ph

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
MEAN	470.05209	1	470.05209	13927.48	.0000
wax	.07008	1	.07008	2.08	.1752
dark	.31008	1	.31008	9.19	.0104
closure	2.85208	1	2.85208	84.51	.0000
wd	.00408	1	.00408	.12	.7340
wc	2.05408	1	2.05408	60.86	.0000
dc	.00075	1	.00075	.02	.8840
wdc	.00675	1	.00675	.20	.6627
ERROR	.40500	12	.03375		

A.8

Paint Film Yellowing

(Levison 1985)

```
# 2V
# The data table is an exact copy of his Table 5 and is not reproduced here.
# An example line:
# 1  5.32  10.17  3.97  6.28  3.69  10.58  4.84  15.66  4.85
#
# Test changes in bleached levels over time.
# Square roots are used because a preliminary run indicated that
# the variance of a group of samples is proportional to the mean.
##############################################################################

/INPUT   var   =  9.
         form  = '(2x, 9f6)'.
         file  = 'levison.data'.
/VAR     name  =  start, dark1,light1, dark2,light2, dark3,light3,
                  dark4,light4.
/TRAN    start =  sqrt(start).
         light1=  sqrt(light1).
         light2=  sqrt(light2).
         light3=  sqrt(light3).
         light4=  sqrt(light4).
/DESIGN depend=  start,  light1,  light2,  light3,  light4.
         level =  5.
         name  =  time.
         orth.
/PRINT   level =  brief.
/END

ANALYSIS OF VARIANCE FOR  - start     light1    light2    light3    light4

SOURCE             SUM OF     DEGREES OF       MEAN           F       TAIL
                   SQUARES     FREEDOM        SQUARE                 PROB.

MEAN              683.57643       1         683.57643      1266.75     .0000
ERROR              16.72853      31            .53963

t(1)                .01013       1            .01013          .53     .4711
ERROR               .58978      31            .01903

t(2)                .05248       1            .05248         9.77     .0038
ERROR               .16656      31            .00537

t(3)                .14837       1            .14837        36.91     .0000
ERROR               .12462      31            .00402

t(4)                .09012       1            .09012        19.54     .0001
ERROR               .14299      31            .00461

time                .30111       4            .07528         9.12     .0000
ERROR              1.02395     124            .00826
```

A.9

Survey Analysis

```
# 1D
# Convert letter codes to number codes and name categories.
# The most important parts of the output for all runs are included in Chapter 5.
#################################################################################################
/INPUT  var   = 14.
        form  = '(a1, i2, a3, 1x, a1, i1, 1x, a2, 1x,
                  3i1, 1x, 2i1, 1x, 3i1)'.
        file  = data.
/VAR    name  = journal, year, id, phase, type, material,
                treatnum, repnum, sampling, table, plot,
                describe, estimate, test.
        label = id.
        add   = new.
/TRANS
journal  = indx(journal, char(A), char(B), char(C), char(D)).
phase    = indx(phase, char(a), char(b), char(c)).
artcat   = 10 * phase + type.
material = indx(material, char(mi), char(fe), char(cu), char(ag), char(mt),
               char(pg), char(cv), char(vn), char(tx), char(dy), char(pp),
               char(wp), char(wd), char(ph), char(dg), char(st), char(cm),
               char(at), char(lt), char(gl), char(iv), char(om), char(ot)).
/GROUP
name(journal) = jaic, studies, techbul, preprint.
code(journal) =    1,       2,       3,       4.
name(year) = year80, year81, year82, year83, year84, year85, year86.
code(year) =     80,     81,     82,     83,     84,     85, 86.
name(phase) = artcomp, artdeter, conserve.
code(phase) =      1,       2,       3.
name(type) = howto, casestud, gensimul, genreal, essay.
code(type) =     1,       2,       3,       4,     6.
name(artcat) =  comphow , compcase, compsim , compreal, compessy,
               detehow , detecase, detesim , detereal, deteessy,
               conshow , conscase, conssim , consreal, consessy,
code(artcat) =  11,       12,       13,       14.       16,
               21,       22.       23.       24.       26,
               31,       32,       33,       34,       36.
name(material) =metal,    metal,    metal,    metal,    metal,    coating,
               substrat, coating,  substrat, coating,  substrat, substrat,
               organic,  coating,  coating,  mineral,  mineral,  other,
               organic,  mineral,  organic,  organic,  other.
code(material) =1,        2,        3,        4,        5,        6,
               7,        8,        9,        10,       11,       12,
               13,       14,       15,       16,       17,       18,
               19,       20,       21,       22,       23.
```

```
# In initial run, used to get number of each material before combine.
#name(material) =marine  , iron    , copper  , silver  , metal   , pigment ,
#               canvas  , varnish , textile , dye     , paper   , wallpap ,
#               wood    , photo   , dagtype , stone   , ceramic , analtech,
#               leather , glass   , ivory   , othermat, othergen.
#code(material) =1,       2,       3,       4,       5,       6,
#               7,       8,       9,       10,      11,      12,
#               13,      14,      15,      16,      17,      18,
#               19,      20,      21,      22,      23.

name(treatnum to test) = inapplic, should, unclear, clear.
code(treatnum to test) =        1,      2,       3,     4.

/PRINT   level = brief.
/SAVE    new.
         file = save.
         code = reedy.
/END


# 8D  correlation of year and statistical variables (treatnum to test) with each other.
#################################################################################################################

/INPUT   file =  save.
         code  = reedy.
/CORR    row   = year, treatnum to test.
         col   = year, treatnum to test.
/PRINT level = brief.       case = 0.
       no freq.
/END

# Repeat for general studies subset by adding the following line.
/TRAN    use   = type eq 3 or type eq 5.
```

```
# 4F
# 1. Frequency tables for pairs of classification variables.
# 2. Log-linear model for all classification variables.
# 3. Percentages for type, journal, and phase.
# 4. Journal, phase, and type versus statistical variables with model.
# 5. Year and phase versus statistical variables with rank correlation.
##############################################################################################################

/INPUT  file = save.    code = reedy.
/TABLE  row  = journal, journal, journal,  year,  year, year,     phase, phase,    type.
        col  = phase,   type,    material, phase, type, material, type,  material, material.
/PRINT  level = brief.  list  = 0.
        no exc.
/END

/INPUT  file = save.    code = reedy.
/TABLE  index = journal, year, phase, type, material.
/FIT    assoc = 3.
/PRINT  level = brief.  list  = 0.
        no exc.
        no obs.
/END

/INPUT  file = save.    code = reedy.
/TABLE  index = type, journal, phase.
/PRINT  level = brief.  list  = 0.
        no exc.
        perc  = tot.
/END

/INPUT  file = save.    code = reedy.
/TRAN   use  = treatnum ne 1.
/TABLE  row  = journal.
        col  = treatnum to test.
        catvar= phase.
        catvar= type.
        cross.
/FIT    assoc = 2.
/PRINT  level = brief.  list  = 0.
        no exc.        no obs.
/END

/INPUT  file = save.    code = reedy.
/TRAN   use  = treatnum ne 1.
/TABLE  row  = year, phase.
        col  = treatnum to test.
        cross.
/STAT   spear.
/PRINT  level = brief.  list  = 0.
        no exc.        obs.
        perc  = row.
/END
```

Glossary

analysis of variance	A technique for measuring the effect of categorical variables on a continuous variable. It is based on dividing (analyzing) the observed variation of the continuous variable into components, which are assigned to the possible effects (see pages 40-44, 58-63).
average	(See "mean").
categorical variable	A variable whose possible values are categories.
category	One of a set of possible values that have no particular ordering. Azurite, lazurite, and cobalt blue are possible values for the variable, "blue pigment."
cluster analysis	A multivariate technique for dividing objects into groups or clusters. (Not seen in the conservation literature, but used in archaeometry and many other fields.)
comparison measures	A quantitative measure of similarity, dissimilarity, or distance between entities as derived from multivariate data.
confidence interval	A numeric interval derived from sample data that expresses our belief about the location of the mean or other measure of a population. The population characteristic is fixed; the interval is variable and depends on the sample. A larger interval lets us be more confident that we have included the true value. (Not seen in the conservation literature, but should be used.)
contingency table	A frequency table with at least two dimensions.
correlation	An observed relationship between two ordered variables such that low and high values of one tend to respectively occur with low and high values of the other (positive correlation) or vice-versa (negative correlation).
cross-tabulation	A frequency table with at least two dimensions.
discriminant analysis	A technique for determining the best way to combine numeric variables to derive a discriminant function that will allow us to assign objects to one of several possible groups or categories. The stepwise version selects a parsimonious subset of the variables. (Not seen in the conservation literature.)
distance measure	A comparison measure varying from zero to infinity that gives a distance between entities. The Euclidean distance based on the Pythagorean formula is only one of many possible distance measures.
estimation	A decision about a value not directly measured based on related information. Regression is one type of estimation.
experimental unit	The entity that receives a particular treatment (see pages 49-51).
F-value	A ratio of two variances used to test a hypothesis as in analysis of variance.

frequency table	A table whose columns represent the categories of a particular variable. If there are multiple lines, each row represents the categories of another variable. The entries in the body of the table are the frequency of occurrence (number of occurrences) of a particular category or combination of categories.
hypothesis test	A decision as to whether observed experimental data are consistent with a particular hypothesis about the system being investigated (see pages 37-38, 58-63).
mean	A summary statistic for numeric variables that indicates where the typical values of a sample or population are located. The arithmetic total of all values divided by their number.
multivariates	Multiple variables measured at the same time and analyzed together. Some multivariate analyses require the same unit of measurement for each variable.
randomization	In its simplest form, the process of selecting entities for measurement or treatment so that each entity has the same probability of being chosen and each is chosen or not independently of the others (see pages 51-53).
regression	The estimation of a functional relationship between one or more variables, often called predictors or independent variables, and a dependent variable (see pages 38-39, 58).
repeated measure	A variable that is measured more than once for each entity in the study (see pages 40-44, 49-52).
replicates	Multiple objects or entities measured under the exact same set of treatment conditions (see pages 49-51).
sampling	The process of choosing which objects to measure when we want to know about a certain class of objects but cannot measure them all (see pages 14-17, 20-22).
scatter plot	A plot in which each point has the corresponding values of the two numeric variables represented by the two axes (Figures 7 and 8).
significant figures	The digits in a number that actually mean something (see page 54).
similarity measure	A measure of resemblence based on a particular set of variables or objects. It usually varies between -1 and 0 or 0 and 1. Correlations measure similarity between variables.
standard deviation	A summary statistic for numeric variables that indicates how much the values of a sample or population are spread away from the mean (see pages 56-57).
standard error	An estimate of the standard deviation of a summary statistic, such as the mean, derived from the standard deviation of a sample (see pages 56-57).
statistic	A number calculated from and summarizing raw data (see page 3).

statistics	Statistics is the art and science encompassing the theory and techniques developed for calculating and using numbers calculated from raw research data. Statistics are used to describe objects, estimate the characteristics of a population from a sample, and test hypotheses or ideas about the subject of a study (see page 3).
t-test	A hypothesis test based on the ratio between a statistic and its standard error (see pages 58-63).
validation	A procedure for establishing that an analytical method really works (see pages 11-13).
variance	A measure of variation; the standard deviation squared.

References

Allesandrini, G., G. Dassu, R. Bugini, and L. Formica
1984 The technical examination and conservation of the portal of St. Aquilano's chapel in the basilica of St. Lorenzo, Milan. *Studies in Conservation* 29(4):161-171.

Allison, Ann H. and Robert B. Pond, Sr.
1983 On copying bronze statuettes. *Journal of the American Institute for Conservation* 23(1):32-46.

Barger, Susan M., A.P. Giri, William B. White, William S. Ginell, and Frank Preusser
1984 Protective surface coatings for daguerreotypes. *Journal of the American Institute for Conservation* 24(1):40-52.

Barger, Susan M., S.V. Krishnaswamy, and R. Messier
1982 The cleaning of daguerreotypes: comparison of cleaning methods. *Journal of the American Institute for Conservation* 22(1):13-24.

Block, Ira
1982 The effect of an alkaline rinse on the aging of cellulosic textiles. Parts I and II. *Journal of the American Institute for Conservation* 22(1):25-36.

Bomford, David and Sarah Staniforth
1981 Wax-resin lining and colour change: an evaluation. *National Gallery Technical Bulletin* 5:58-65.

Bowman, Janet Gilliland and Barbara M. Reagan
1983 Filtered and unfiltered lights and their effects on selected dyed textiles. *Studies in Conservation* 28(1):36-44.

Branchick, Thomas J., Keiko M. Keyes, and F. Christopher Tahk
1982 A study of the bleaching of naturally aged paper by artificial and natural light. *AIC Preprints* 1982:29-39.

Brill, Robert H., I. Lynus Barnes, and Thomas J. Murphy
1981 Lead isotope studies of the bronze horse from the Metropolitan Museum of Art. Appendix IV in "Technical examination of the classical bronze horse from the Metropolitan Museum of Art" by Kate C. Lefferts, Lawrence J. Majewski, Edward V. Sayre, and Pieter Meyers. *Journal of the American Institute for Conservation* 21(1):32-39.

Butler, Marigene H.
1984 An investigation of the materials and techniques used by Paul Cezanne. *AIC Preprints*:20-33.

Calamiotou, M., M. Siganidou, and S. E. Filippakis
1983 X-ray analysis of pigments from Pella, Greece. *Studies in Conservation* 28(3):117-121.

Clapp, Anne F.
1981 The examination of Winterthur wallpapers and a progress report, April 1980, on a group of papers from the Fisher House, Philadelphia. *Journal of the American Institute for Conservation* 20(2):66-73.

Clement, Daniel
1983 The blistering of paper during hydrogen peroxide bleaching. *Journal of the American Institute for Conservation* 23(1):47-62.

Cordy, Ann and Kwan-nan Yeh
1984 Blue dye identification of cellulosic fibers: indigo, logwood, and Prussian blue. *Journal of the American Institute for Conservation* 24(1):33-39.

Crews, Patricia Cox
1982 The influence of mordant on the lightfastness of yellow natural dyes. *Journal of the American Institute for Conservation* 21(2):43-58.

de la Rie, E. René
1982 Fluorescence of paint and varnish layers (Part III). *Studies in Conservation* 27(3):102-108.

DeSantis, Pia C.
1983 Some observations on the use of enzymes in paper conservation. *Journal of the American Institute for Conservation* 23(1):7-27.

Dixon, W. J. (editor)
1985 *BMDP Statistical Software.* Berkeley: UC Press.

England, P.A. and L. van Zelst
1982 A technical investigation of some seventeenth-century New England portrait paintings. *AIC Preprints*:85-95.

Fiedler, Inge
1984 Materials used in Seurat's *La Grande Jatte,* including color changes and notes on the evolution of the artist's palette. *AIC Preprints*:43-51.

Gifford, E. Melanie
1983 A technical investigation of some Dutch 17th century tonal landscapes. *AIC Preprints*:39-49.

Gilberg, Mark R. and Nigel J. Seeley
1982 Liquid ammonia as a solvent and reagent in conservation. *Studies in Conservation* 27(1):38-44.

Gilmore, Andrea M.
1981 Wallpaper and its conservation—an architectural conservator's perspective. *Journal of the American Institute for Conservation* 20(2):74-82.

Hackney, S. and G. Hedley
1981 Measurements of the ageing of linen canvas. *Studies in Conservation* 26(1):1-14.

Indictor, N., R.J. Koestler, and R. Sheryll
1985 The detection of metallic mordants by energy dispersive X-ray spectrometry. Part I. Dyed woolen textile fibers. *Journal of the American Institute for Conservation* 24(2):104-109.

Johnston-Feller, Ruth, Robert L. Feller, Catherine W. Bailie, and Mary Curran
1984 The kinetics of fading: opaque paint films pigmented with alizarin lake and titanium dioxide. *Journal of the American Institute for Conservation* 23(2):114-129.

Katz, Kenneth B.
1985 The quantitative testing and comparison of peel and lap/shear for Lascaux 360 H.V. and BEVA 371. *Journal of the American Institute for Conservation* 24(2):60-68.

Koestler, R.J., N. Indictor, and R. Sheryll
1985 The detection of metallic mordants by energy dispersive X-ray spectrometry. Part II. Historical silk textiles. *Journal of the American Institute for Conservation* 24(2):110-115.

Lefferts, Kate C., Lawrence J. Majewski, Edward V. Sayre, and Pieter Meyers
1981 Technical examination of the classical bronze horse from the Metropolitan Museum of Art. *Journal of the American Institute for Conservation* 21(1):1-42.

Levison, Henry W.
1985 Yellowing and bleaching of paint films. *Journal of the American Institute for Conservation* 24(2):69-76.

Marchese, B. and V. Garzillo
1984 An investigation of the mosaics in the cathedral of Salerno. Part II: Characterization of some mosaic tesserae. *Studies in Conservation* 29(1):10-16.

McClintock, Thomas K.
1981 The *in situ* treatment of the wallpaper in the study of the Longfellow National Historic Site. *Journal of the American Institute for Conservation* 20(2):111-115.

Nelson, J., A. King, N. Indictor, and D. Cabelli
1982 Effects of wash water quality on the physical properties of three papers. *Journal of the American Institute for Conservation* 21(2):59-76.

Newman, Richard and Gridley McKim-Smith
1982 Observations on the materials and painting technique of Diego Velazquez. *AIC Preprints*:133-140.

Nosek, Elisabeth M.
1985 The investigation and conservation of a lead paten from the eleventh century. *Studies in Conservation* 30(1):19-22.

Orna, Mary Virginia and Thomas F. Mathews
1981 Pigment analysis of the Glajor Gospel book of UCLA. *Studies in Conservation* 26(2):57-72.

Parrent, James M.
1985 The conservation of waterlogged wood using sucrose. *Studies in Conservation* 30(2):63-72.

Peacock, Elizabeth E.
1983 Deacidification of degraded linen. *Studies in Conservation* 28(1):8-14.

Pearlstein, E. J., D. Cabelli, A. King, and N. Indictor
1982 Effects of eraser treatment on paper. *Journal of the American Institute for Conservation* 22(1):1-12.

Phillips, Morgan W.
1984 Notes on a method for consolidating leather. *Journal of the American Institute for Conservation* 24(1):53-56.

Reagan, Barbara
1982 Eradication of insects from wool textiles. *Journal of the American Institute for Conservation* 21(2):1-34.

Rodriguez, J.L., C. Maqueda, and A. Justo
1985 A scientific study of the terracotta sculptures from the porticos of Seville Cathedral. *Studies in Conservation* 30(1):31-38.

Roy, Ashok
1982 Hogarth's *Marriage à la Mode* and contemporary painting practice. *National Gallery Technical Bulletin* 6:59-67.

Sack, Susanne P., F. Christopher Tahk, and Theodore Peters, Jr.
1981 A technical examination of an ancient Egyptian painting on canvas. *Studies in Conservation* 26(1):15-23.

Science (editors)
1987 Information for contributors. *Science* 235 (March 27):xi.

Simunkova, E., J. Brothankova-Bucifalova, and J. Zelinger
1985 The influence of cobalt blue pigments on the drying of linseed oil. *Studies in Conservation* 30(4):161-166.

Simunkova, E., Z. Smejkalova, and J. Zelinger
1983 Consolidation of wood by the method of monomer polymerization in the object. *Studies in Conservation* 28(3):133-144.

Skoulikidis, Theodore N. and Nicholas Beloyannis
1984 Inversion of marble sulfation-reconversion of gypsum films into calcite on the surfaces of monuments and statues. *Studies in Conservation* 29(4):197-204.

Stodulski, L., E. Farrell, and R. Newman
1984 Identification of ancient Persian pigments from Persepolis and Pasargadae. *Studies in Conservation* 29(3):143-154.

Wang, Y. and A. P. Schniewind
1985 Consolidation of deteriorated wood with soluble resins. *Journal of the American Institute for Conservation* 24(2):77-91.

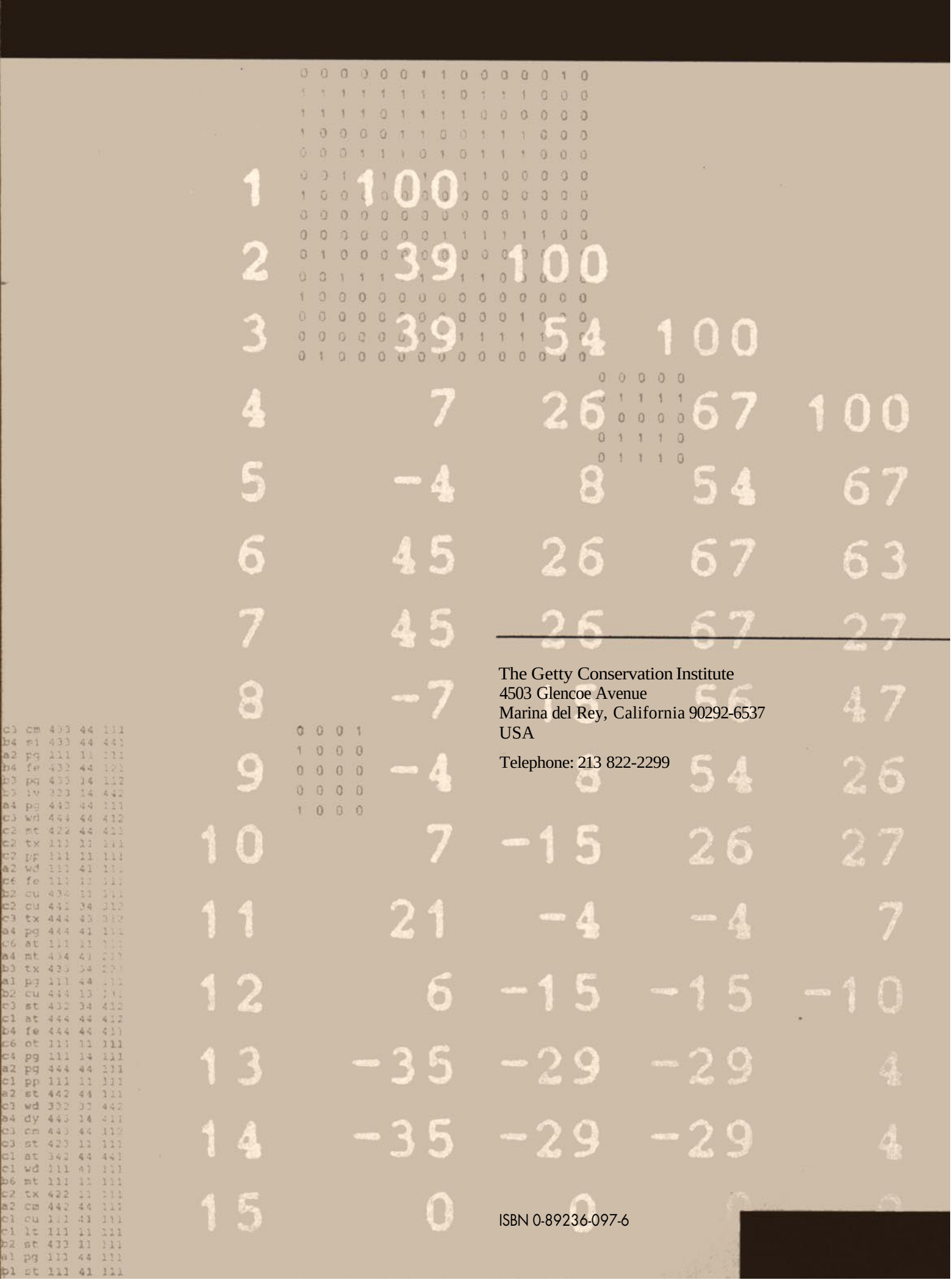
Whitmore, Paul M., Glen R. Cass, and James R. Druzik
1986 The fading of traditional natural colorants due to atmospheric ozone. *AIC Preprints*:114-124.

Wilhelm, Henry
1981 Monitoring the fading and staining of color photographic prints. *Journal of the American Institute for Conservation* 21(1):49-64.

Winter, John
1981 'Lead white' in Japanese paintings. *Studies in Conservation* 26(3):89-101.

Wouters, Jan
1985 High performance liquid chromatography of Anthraquinones--analysis of plant and insect extracts and dyed textiles. *Studies in Conservation* 30(3):119-128.

Index



The Getty Conservation Institute
4503 Glencoe Avenue
Marina del Rey, California 90292-6537
USA

Telephone: 213 822-2299

ISBN 0-89236-097-6