Alternative Models for Gauge Studies

SEMATECH
Technology Transfer 93081755A-TR

Alternative Models for Gauge Studies

Technology Transfer # **93081755A-TR**
SEMATECH
February 24, 1994

Abstract: This document describes the results of a study of gauge experiments performed from the summer 1992-winter 1993 and is based on a series of reports prepared monthly during the study. Classical gauge study—the subdivision of the sum of squares between sites to investigate the bull's-eye phenomenon—the analysis of the general factorial experiment, and the problems of missing data and unbalanced experiments are discussed. The final section suggests a generic design for gauge studies on wafers.

Keywords: Metrology, Wafers, Gauge Studies

Approvals: Dr. Peter John, Ph.D., Author
Paul Tobias, Manager
Ed Hall, Director
Susan Rogers, Manager, Technical Information Transfer
Richard C. Johnson, Director, Technology Transfer
Jeanne Cranford, Technical Information Transfer Team Leader

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am indebted to my colleagues Veronica Czitrom, Arnon Hurwitz, Kyle Merrill, Jack Reece, and Pat Spagon at SEMATECH for helpful discussions during the period when the study was conducted. I am particularly grateful to Mark Sorell who presented some of this material to a SEMATECH semi-annual seminar during my absence. His presentation stimulated helpful discussion and comments from several members of supporting companies. Some of these comments have been incorporated in the revised version of this report. To have incorporated all of the comments would have necessitated a document twice or three times the size of this one and could not have been written within the time allotted for this project.

Preface

Gauge studies is a large field that depends heavily on factorial experiments to estimate variance components. This subject has been associated for years with agriculture, but engineers only lately have been concerned with more than the most elementary aspects.

Today's engineers must measure at very high precision using equipment costing hundreds of thousands of dollars. Variablity among measuring devices is no longer just a nuisance; it has become a matter of critical importance. To put it bluntly, engineers are called to answer the question "Is this expensive measurement machine up to the task it is being purchased for?

Another characteristic of measurement devices is that they are highly automated. In theory there is little variability between operators or their training (as was the case in the past) which can lead to different interpretations of a chemical test. One question not considered in the study was if the measuring device measuring the correct response. A device that measures thickness within a 10% bias but with low repeatability may not be what the engineer really wants.

The concepts of repeatability and reproducibility are inherited from chemistry. Repeatability presents little difficulty. For modern measuring instruments, it contributes only a very small fraction of the variability in measurements. On the other hand, the definition of reproducibility is difficult to define precisely because there seems to be little general agreement on the details. The basic idea of reproducibility is clear enough, but uncertainty arises from its definition as a linear combination of the components of variance. Different engineers include different components in their definitions. This disagreement over the components of variance can lead to confusion and to frustrating dialogues among engineers.

The author used MINITAB software to make calculations throughout this document. This package is the one that I use in my own teaching and research. I do not deal with large data sets; consequently, my own knowledge of SAS and RS1 is limited. The calculations here can also be made routinely with those packages.

1 EXECUTIVE SUMMARY

This document describes the results of a study of gauge experiments performed from the summer 1992-winter 1993 and is based on a series of reports prepared monthly during the study. Classical gauge study, the subdivision of the sum of squares between sites to investigate the bull's-eye phenomenon, the analysis of the general factorial experiment, and the problems of missing data and unbalanced experiments are discussed. The final section suggests a generic design for gauge studies on wafers.

2 THE CLASSICAL GAUGE STUDY

This project began with an examination of a classical repeatability and reproducibility (R and R) study. In its traditional form, R and R is no longer appropriate to our work because of the increased complexity of modern metrology and the impact of automation in reducing operator variability. But the examination of this procedure provided a starting point for the investigation.

One of the difficulties in this study is the numerous conflicting definitions of *repeatability* and *reproducibility,* and some of the definitions are not appropriate to our applications. Four descriptions of the classical procedure can be referred to (the full citations are listed in Section 9, References). They are given by:

1. Barrentine, Larry B. (1991),
2. Montgomery, Douglas C. and George Runger (1992),
3. Wheeler, Donald J. and Richard W. Lyday (1989),
4. Wheeler, Donald J. (1992)

The topic, in its modern context, is so new to most engineers that authors do not even agree on the spelling of the key word. Should it be gage or gauge? The latter spelling has become the standard usage.

All four authors cited write about using sample ranges to obtain estimates of standard deviations. That practice is a relic of classical quality control. There is no need to still follow that outdated procedure. Models with the analysis of variance should be used. Software packages are readily available.

2.1 The Standard Experiment

Each author considers one or more operators. In the standard example, each operator makes two observations (replicates) on each of several similar parts. The repeatability of the measuring instrument is a function of the variance between the duplicate observations made by the same operator on the same part. It is convenient to denote this variance by σ^2_r. The σ_r is estimated by taking the average of the ranges of each operator on each part and multiplying by the appropriate factor ($1/d_2$) from the standard tables. If there are o operators and p parts, there will be op ranges to be averaged.

Barrentine [1] bases his explanation of repeatability upon the General Motors Long Form. Both he and Wheeler [3] refer to the standard deviation, σ_r, rather than the variance as the repeatability. The Long Form then compares it to the specified tolerance by the ratio:

$$5.15 * \sigma_r / \text{spec.tolerance}$$

which Barrentine [1] calls the equipment variation (EV). Wheeler [3] concurs that multiplying σ_r by 5.15 has been the traditional approach, but he further argues that the comparison does not work. His argument is that the use of the multiplier gives unreasonable answers. That practice will not be followed in this study.

On the other hand, Montgomery [4] denotes σ^2_r by σ^2_{gauge} and takes the ratio P/T (precision to tolerance) defined by:

$$P/T = 6\sigma_{gauge} / (USL - LSL).$$

Although there is a lack of standardization in this field about even the basic notations and nomenclature, there is general agreement that repeatability is a measure of the variability of measurements at the same time and under the same conditions.

Reproducibility suffers more than repeatability from the problem of definition. Wheeler [3] takes the range of operator averages and divides that by the appropriate value of d_2 to get an estimate of the standard deviation between operators. Barrentine [1] adjusts for the repeatability and multiplies by 5.15. Wheeler [3] then squares both standard deviations, adds them, and calls the square root R and R. He then adds the variance between parts to obtain a measure of the *total variation.* This differs in terminology but not in basic definition from Montgomery [4] who defines total variation as:

$$\sigma^2_{total} = \sigma^2_{product} + \sigma^2_{gauge}$$

2.2 An Example From Wheeler

The data for this example, in Table 2.2.1 was presented by Wheeler [3].

Table 2.2.1 Gasket Thickness

Part	Operator		
	A	**B**	**C**
1	67	55	52
	62	57	55
2	110	106	106
	113	99	103
3	87	82	80
	83	79	81
4	89	84	80
	96	78	82
5	56	43	46
	47	42	54

The repeatability is calculated from the cell ranges. The average range is given by the following formula:

$$\overline{R} = \left[(67-62)+(113-110)+\ldots+(54-46)\right]/15 = 4.2667$$

The repeatability standard deviation is obtained by dividing the average range by the appropriate factor (d_2 = 1.128). This gives:

$$\sigma_r = 4.2667/1.128 = 3.783$$

It is more convenient to work in terms of the variance, σ_r^2 = 14.31. The repeatability will be defined as the variance between duplicate observations made under identical conditions.

There is an implied mathematical model: let y_{ijk} denote the k^{th} observation by the i^{th} operator on the j^{th} part, then:

$$y_{ijk} = \mu + o_i + p_j + e_{k(ij)} \qquad (2.2.1)$$

In this model, μ is a constant term, o_i is the contribution of the i^{th} operator, p_j is the contribution of the j^{th} part, and $e_{k(ij)}$ is the repeatability error. The term o_i is a random variable that is assumed to be normally distributed with mean zero and variance σ_o^2; $e_{k(ij)}$ is normally distributed with mean zero and variance σ_r^2; and o_i and $e_{k(ij)}$ are independent of each other.

The assumption of normality is required when one uses the range method to obtain estimates because divisors such as d_2 depend upon this type of distribution. The same assumption is not essential when using the analysis of variance approach unless one wishes to go further and use tests of significance. Even then one has a certain degree of protection from the central limit theorem.

The averages for the three operators are A = 81.0, B = 72.5, C = 73.9. Their range, 8.5, is divided by the value of d_2 for a sample of three observations to give an estimate of the reproducibility standard deviation:

$$\sigma_R = \sigma_o = 8.50/1.693 = 5.021$$

Note: Wheeler [3] points out that this estimate is biased. The unbiased estimate is given by:

$$\sigma_o^2 = 5.021^2 - \sigma_r^2/10 = 23.78$$

where 10 is the number of observations made by each operator.

The ANOVA procedure with the model in the equation Section 2.2.1 gives different estimates:

$$\hat{\sigma}_r^2 = 12.45, \hat{\sigma}_o^2 = 19.53$$

The ANOVA procedure gives estimates that are based upon a least squares partition of the total variation that corresponds to maximum likelihood estimates in the case of normality. These estimates are more precise than those obtained by the range method.

Note that Wheeler's [3] implied model does not include a part x operator interaction term. The inclusion of that term makes modest changes in the estimates:

$$\hat{\sigma}_r^2 = 12.20, \hat{\sigma}_o^2 = 19.48$$

but the F value for the interaction is only 1.06, which justifies its omission.

2.3 An Alternative Point of View

In presenting the traditional method of analysis in Section 2.2, using the analysis of variance rather than the range method to obtain estimates of the repeatability and reproducibility variances was recommended. The recommendation is not based only on mathematical niceties, nor on the fact that the two methods give different estimates.

This report will take the point of view that using the ranges is inappropriate to modern metrology in the semiconductor industry. It will be suggested that the model in equation 2.2.1 is inadequate and that the emphasis on operator differences as the source of reproducibility variance is not the correct approach.

2.4 Two Variations on the Basic Experiment

How are these approaches adapted to the study? The first tentative steps suggested two possible experiments. In Section 2.5 the basic experiment that incorporates some of the major lessons learned during the study will be presented.

2.5 Metrology

An operator loads a wafer into a measuring instrument, or machine, and makes observations on a response (e.g., thickness) at several sites. The number of sites is denoted by s, which may be 5, 9, or 49. The same sites are used repeatedly; consequently, sites are treated as a fixed effect.

2.6 Experiment 1

A single operator takes a wafer, loads it in the machine and makes n measurements at each site. The measurements are *not* taken at random but in cycles, and it is assumed that there is no difference between the cycles. The machine makes a measurement at each site in sequential order. Then it makes a second cycle and takes another measurement. The data table has s columns (sites) with n observations in each column. A one-way ANOVA is done in which two sums of squares are obtained: between sites and within sites. The sum of squares within sites, which is also the mean square for error, estimates the repeatability variance σ_r^2.

In the mathematical model let y_{ij} denote the j^{th} measurement at the i^{th} site in the following model:

$$y_{ij} = \mu + s_i + e_{ij};$$

s_i denotes the i^{th} site, e_{ij} is the repeatability term.

This experiment is not adequate. It addresses only the repeatability and ignores any variability in the loading of the wafer, which is the main way the operator can influence the measuring process.

2.7 Experiment 2

Experiment 2 differs from experiment 1 in that the operator stops the machine after each cycle, removes the wafer, loads it in the machine again, and then makes the next cycle.

Now there is a two-way layout, sites compared to cycles or loadings. Cycles and loadings are identical. There is one observation per cell; and loadings is a random effect. With only one observation per cell the repeatability is estimated by the mean square for site x loading interaction. The mean square between loadings provides an estimate of the variance due to loadings.

The model is changed to:

$$y_{ij} = \mu + s_i + p_j + e_{ij};$$

p_j is the contribution of the j^{th} loading; it is a random variable with variance σ_p^2.

Note: It is more convenient when typing to use the letter p to denote loading rather than the letter l. The latter can too easily be confused with the number 1. Therefore, the term *placement* rather than *loading* will be used in this report.

The expectation of the mean square for placements is $\sigma_r^2 + s\sigma_p^2$; and the expectation of the error, or SxP, mean square is σ_r^2. Estimates of σ_p^2 can be calculated and called the reproducibility. The total measurement error of the measuring system is:

$$\sigma_T^2 = \sigma_r^2 + \sigma_p^2$$

As additional sources of variation are introduced, the problem becomes much more complex. Is there drift in the machine over time? Is there interaction between sites and placements? Between sites and cycles? How are placements and cycles separated? What is the appropriate model for calculating the expected mean squares in ANOVA calculations with mixed models? Some of these questions will be discussed in the sections that follow.

3 GAUGE EXPERIMENTS WITH WAFERS

A more specialized scenario for the basic experiment is now considered and referred to as experiment A. A measuring device requires the operator to place a wafer in the device and then press a button. After that action is taken, the device automatically makes as many repeat readings as is desired.

As in the earlier experiment, o operators are taken. The operators may be a random sample from a population of operators, or they may be Art and Bill, the only two operators in the shop—a fixed factor. For the moment it is assumed the operators are a random sample from some population of operators. A random sample of w wafers is also taken. Each operator tests each wafer by placing it in the device, pressing the button and recording r readings (observations) for a total of N = owr observations.

This is a crossed experiment since wafers are crossed with operators and each operator measures every wafer. There are ow cells and each cell contains r observations. There is also a split-plot facet to the experiment. Each cell corresponds to a placement of a wafer in the device by an operator. Observations in the same cell are not independent because they are made with the same (random) placement.

There are four sources of variability in the observations:

1. Between wafers with variance σ_w^2.
2. Between operators with variance σ_o^2.
3. Between placements with variance σ_p^2.
4. Between repeat observations on the same wafer with the same placement with variance σ_r^2.

The first component, σ_w^2, is of no interest in the metrology experiment because it is not attributable to the measuring process. The balance built into the experiment enables the investigators to eliminate it from the other components in the ANOVA procedure.

The fourth component, σ_r^2, is the same as in sections 2.6 and 2.7, the repeatability. It is the variance within cells in the ANOVA; and is expected to be very small.

The second and third components, σ_o^2 and σ_p^2 appear in the reproducibility variance, which is:

$$\sigma_R^2 = \sigma_o^2 + \sigma_p^2$$

3.1 The Mathematical Model

There are ow cells (placements) arranged in o columns (operators) and w rows (wafers); and there are r observations in each cell. Let y_{ijk} denote the k^th observation in the (ij)^th cell:

$$y_{ijk} = \mu + o_i + w_j + p_{ij} + e_{k(ij)}$$

where o_i, w_j, p_{ij} and $e_{k(ij)}$ are independent (normal) random variables with zero means and variances σ_o^2, σ_w^2, σ_p^2, and σ_r^2 respectively.

The standard analysis of variance procedure divides the total sum of squares into four constituent sums. They are listed in Table 3.1.1 together with their expected mean squares.

Table 3.1.1 ANOVA

Source	DF	EMS
wafers	w-1	$\sigma_r^2 + r\sigma_p^2 + or\sigma_w^2$
operators	o-1	$\sigma_r^2 + r\sigma_p^2 + wr\sigma_o^2$
WxO = placements	(o-1)(w-1)	$\sigma_r^2 + r\sigma_p^2$
within cells	ow(r-1)	σ_r^2

It is important to realize that in the ANOVA calculations the sum of squares for placements corresponds to the WxO interaction sum of squares. More will be said about this in Section 8.

3.2 Reproducibility

It is argued that if a particular wafer was given to an operator chosen at random who then placed it in the device and then pushed the button and recorded an observation y, that observation would have variance:

$$V(y) = \sigma_r^2 + \sigma_p^2 + \sigma_o^2$$

The first term is the repeatability. The sum of the last two terms is the reproducibility. In this experiment, the repeatability is estimated by the mean square within cells or M_e. To estimate the reproducibility we equate the mean squares for operators and for operator x wafer interaction to their expected values and solve for σ_p^2 and σ_o^2.

The concept of variance between operators raises questions. The essential idea of a difference between two operators is that the values of o_i are different. If, for example, $o_1 - o_2 = 1$, then, apart from random noise, Art will get readings on any wafer that are one point higher than Bill's. It could be said that there is a systematic difference between the two men. Wheeler and Lyday [2] call it operator bias effect. Such a difference is understandable in older devices in which the operator himself makes the reading and uses judgement based on experience and training. Just when has the endpoint in a given test been reached? Some operators may be taught to interpret it in a slightly but systematically different way from one another or to employ a slightly different technique. The differences in observed values could even come from physical differences between the operators involving traits such as hearing, when taking blood pressure readings, or the ability of the eyes to detect color changes.

These problems were important in classical metrology. One can, however, argue that in modern measuring devices, since the readings are made automatically, systematic differences between operators have been eliminated. Negligible variability due to operators should be expected.

On the other hand, it would not be surprising if σ_p^2 were significant. An operator might not be able to place the wafer in precisely the same position each time. That is a source of variability that is properly attributable to the measuring system.

3.3 An Example

In this example, six wafers are tested by three operators. Each operator places each wafer in the machine and then takes two duplicate readings. The data is shown in Table 3.3.1. The ANOVA table is Table 3.3.2.

Table 3.3.1 Data

Operators	Wafers					
	1	**2**	**3**	**4**	**5**	**6**
A	70.52	74.40	73.54	75.20	73.99	71.89
	71.03	76.24	74.68	74.33	75.39	74.70
B	71.08	75.78	70.93	77.40	75.04	72.98
	71.89	76.90	72.93	75.90	74.01	69.87
C	70.56	76.88	75.34	75.65	73.91	73.73
	68.61	78.61	76.35	75.02	75.46	71.55

Table 3.3.2 Analysis of Variance Table
MINITAB Printout

```
MTB >      NAME c1 'Y'  c2 'WAFER' c3 'OPER'
MTB >      ANOVA Y=WAFER OPER WAFER*OPER;
SUBC>      RANDOM WAFER OPER;
SUBC>      EMS.

Analysis of variance for Y

Source               DF          SS         MS         F          P
WAFER                 5     137.034     27.407      9.11      0.002
OPER                  2       2.307      1.154      0.38      0.691
WAFER*OPER           10      30.078      3.008      2.17      0.073
=PLACEMENT
REPEATABILITY        18      24.910      1.384
Total                35     194.329

Source     Variance Error Expected Mean Square
           Component Term (Using Unrestricted Model)
1 WAFER              4.0665    3       (4) + 2(3) + 6(1)
2 OPER              -0.1545    3       (4) + 2(3) + 12(2)
3 PLACEMENT          0.8120    4       (4) + 2(3)
4 REPEATABILITY      1.3839            (4)
                                MEANS
                         _                                         _
WAFER      N             y                OPER       N             y
  1        6        70.615                  1        12       73.826
  2        6        76.468                  2        12       73.726
  3        6        73.962                  3        12       74.306
  4        6        75.583
  5        6        74.633
  6        6        72.453
```

Considering Table 3.3.2, these observations can be made. The first command, preceded by MTB >, states the model that has two factors, WAFER and OPER, and the model includes their main effects and interaction. The first subcommand says that WAFER and OPER are both random factors. The second subcommand calls for printing the expected mean squares.

The line: 1 WAFER 4.0665 3 (4) + 2(3) + 6(1) says that the estimate of σ_w^2 is 4.0665, that the appropriate denominator for the F test for the hypothesis $\sigma_w^2 = 0$ is the mean square in line 3 of the ANOVA table, and that the expected value of the mean square between wafers is:

$$E(M_w) = \sigma_r^2 + 2\sigma_p^2 + 6\sigma_w^2$$

The F statistic for testing $\sigma_w^2 = 0$ is 27.407 / 3.008 = 9.11. This is compared to the tabulated value of the F statistic with 5 d.f. in the numerator and 10 d.f. in the denominator. The observed ratio, 9.11, has a P value 0.002. Thus the hypothesis would be rejected and the variance would be declared significant even if so conservative a value as 0.002 were to be used for α.

The estimate of σ_p^2 is obtained by subtracting the mean square for repeatability from the mean square for placements and dividing by two: (3.008 - 1.384)/2 = 0.8120.

Similarly σ_w^2 is estimated by (27.407 - 3.008) /6 = 4.0665 and σ_o^2 by (1.154 - 3.008) /12 = - 0.1545. There can be no negative variance so it is concluded that $\sigma_o^2 = 0$. A similar thing could happen in the example of Section 4.2 and the reader is referred Section 4.3 for some comments on this problem.

Satisfied that there is no significant difference between operators, the experimenter might decide to pool the sums of squares for operators with the sum of squares for placement. The revised ANOVA table is shown in Table 3.3.3.

Table 3.3.3 Revised ANOVA Table Omitting the Term for Operators

```
Source              DF          SS        MS        F
WAFER                5        137.034   27.407    10.01
PLACEMENT           12         32.385    2.738
REPEATABILITY       18         24.910    1.384
Total               35        194.329


Source     Variance Error Expected Mean Square
           Component Term (Using Unrestricted Model)

1 WAFER               4.115      2       (3) + 2(2) + 6(1)
3 PLACEMENT           0.677      3       (3) + 2(2)
4 Error               1.384              (3)
```

3.4 A Caution

In the experiment above, the operator made one placement for each wafer and then made two duplicate readings. The engineer might be tempted to make two separate placements for each wafer with one reading at each placement. That approach is not correct because there would be no duplicate observations made under *identical* conditions. No valid estimate of the repeatability or of the variance due to placement could be obtained. The expected value of the mean square for error would now become $\sigma_r^2 + \sigma_p^2$ and the two components could not be separated.

4 A NESTED EXPERIMENT

In experiment A, each operator tested the same wafers. What would happen if a different set of wafers was assigned to each operator? The factor "wafers" would no longer be crossed with operators. Instead it would be nested within operators. Since the operators are not considered important, they are no longer included in the scenario. They are replaced by another random factor—"batches".

4.1 Experiment B—a Nested Experiment

Suppose b batches of wafers are made, w wafers are taken from each batch, and each wafer is placed in the machine on p occasions. On each placement, r repeat observations are made.

The design chosen is called a nested design. The wafers are nested within their batches, the placements are nested within their wafers, and the observations are nested within placements. The use of parentheses in the subscripts in the following model will indicate that nesting. These designs are also called hierarchic designs because observations, placements, wafers, and batches can be considered as steps in a hierarchy, similar to a geneological tree structure.

Each observation is a random variable. It is subject to four sources of variation:

1. Variation between batches, denoted by σ_b^2.
2. Variation between wafers (from the same batch).
3. Variability between placements.
4. The repeatability variance.

A mathematical model for a typical observation is written. Let y_{hijk} denote the k^{th} reading at the j^{th} placement for the i^{th} wafer from the h^{th} batch. Then:

$$y_{hijk} = \mu + b_h + w_{i(h)} + p_{j(ih)} + e_{k(hij)} \qquad (4.1.1)$$

where $1 \le h \le b$, $1 \le i \le w$, $1 \le j \le p$, $1 \le k \le r$.

In this model μ is the "true" value for the process; b_h represents the h^{th} batch; $w_{i(h)}$ represents the wafer; the nesting is indicated by having the subscript in parentheses; $p_{j(hi)}$ is the term for the placement; and $e_{k(hij)}$ is the repeatability error in the observation. Each term, except μ, is a *random variable*. On the average each term is zero. The random variables are mutually independent and their variances are, respectively, σ_b^2, σ_w^2, σ_p^2, σ_r^2.

Thus the variance of an observation is:

$$V(y_{hijk}) = \sigma_b^2 + \sigma_w^2 + \sigma_p^2 + \sigma_r^2 \qquad (4.1.2)$$

Again, the primary purpose of the experiment is to obtain estimates of these variances.

4.2 An example

This example has four batches, five wafers per batch, three placements per wafer, and two observations per placement. Only two observations per placement have been included as a matter of editorial convenience; and in practice r would often be larger. The data are shown in Table 4.2.1. The analysis of variance table is given in Table 4.2.2.

Table 4.2.1 Data

Batch	Wafer	Placement 1		Placement 2		Placement 3	
		Observation 1	Observation 2	Observation 1	Observation 2	Observation 1	Observation 2
1	1	49.26	49.35	50.71	49.57	48.51	49.24
	2	52.06	50.39	44.64	48.76	50.91	49.42
	3	49.30	51.18	49.96	52.01	49.42	49.37
	4	51.89	52.28	51.00	50.80	52.21	51.45
	5	48.09	48.61	47.19	45.19	47.31	46.89
2	6	48.33	49.25	47.20	47.81	45.25	47.14
	7	46.15	47.21	49.71	49.33	48.22	49.50
	8	50.69	50.08	51.49	52.08	46.21	48.33
	9	50.33	48.89	49.91	49.34	50.11	49.00
	10	49.73	47.97	47.26	47.15	48.58	49.04
3	11	51.09	51.85	55.95	56.34	53.70	54.08
	12	52.85	52.09	52.40	52.35	53.42	51.37
	13	50.68	49.98	52.28	54.97	51.90	51.90
	14	52.91	54.49	53.38	54.16	57.13	55.40
	15	48.52	47.90	50.21	50.03	51.78	51.14
4	16	49.82	50.07	50.75	50.40	53.37	53.19
	17	51.33	49.79	46.50	49.97	46.13	46.92
	18	52.63	52.72	51.46	52.26	50.85	49.84
	19	53.43	50.51	47.79	49.02	54.26	52.58
	20	55.07	52.01	50.48	46.69	53.44	54.03

Table 4.2.2 Analysis of Variance Table MINITAB Printout
MINITAB Printout

```
MTB >         ANOVA Y=B W(B) P(B W);
SUBC>         RANDOM B W P;
SUBC>         EMS.

Analysis of Variance for Y

Source   DF   SS        MS       F          P
B        3    251.894   83.965   6.42       0.005
W(B)     16   209.418   13.089   2.37       0.014
P(B W)   40   220.537    5.513   5.06       0.000
Error    60    65.405    1.090
Total    119  747.254

Source   Variance Error Expected Mean Square
             Component  Term
 1 B         2.363    2       (4) + 2(3) + 6(2) + 30(1)
 2 W(B)      1.263    3       (4) + 2(3) + 6(2)
 3 P(B W)    2.212    4       (4) + 2(3)
 4 Error     1.090            (4)
```

The following observations can be made concerning Table 4.2.2. The first command states the model: wafers (W) are nested within batches (B); placements (P) are nested in batches and wafers. The first subcommand says that B, W, P are all random factors. The second subcommand calls for printing the expected mean squares.

The line: 1 B 2.363 2 (4) + 2(3) + 6(2) + 30(1) says that the estimate of σ_b^2 is 2.363, that the appropriate denominator for the F test for the hypothesis $\sigma_b^2 = 0$ is the mean square in line 2 and that the expected value of the mean square between batches is

$$E(M_b) = \sigma_e^2 + 2\sigma_p^2 + 6\sigma_w^2 + 30\sigma_b^2.$$

The F statistic for testing $\sigma_b^2 = 0$ is 83.965/13.089 = 6.42. This is compared to the F statistic with b-1 = 3 d.f. in the numerator and b(w-1) = 16 d.f. in the denominator.

The estimate of σ_b^2 is obtained by subtracting the mean square for wafers from the mean square for batches and dividing by 30: (83.965 - 13.089) / 30 = 2.363.

In a similar way, σ_w^2 is estimated by (13.089 - 5.513)/6 = 1.263 and σ_p^2 by (5.513-1.090)/2 = 2.212.

4.3 Estimating the Variance Components

Let N = bwpr denote the total number of observations.

Consider the r observations, $y_{bwp1}, \ldots, y_{bwpr}$, made on the last of the bwp placements and let their average be $y_{bwp.}$. Then

$$y_{bwpk} - y_{bwp.} = e_{bwpk} - e_{bwp.}$$

and

$$E[\Sigma_k(y_{bwpk} - y_{bwp.})^2] = (r-1)\sigma_r^2$$

Repeating for all the placements and summing results in:

$$E[\Sigma_h\Sigma_i\Sigma_j\Sigma_k(y_{hijk} - y_{hij.})^2] = bwp(r-1)\sigma_r^2 \qquad (4.3.1)$$

The sum of squares on the left can be called the sum of squares between duplicate readings (on the same placement) denoted by S_r. It simplifies to:

$$S_r = \Sigma_h\Sigma_i\Sigma_j(\Sigma_k y_{hijk}^2 - r\bar{y}_{hij.}^2)$$

or, if T_{hij} is written for the total of the observations on the $(hij)^{th}$ placement,

$$S_r = \Sigma_h\Sigma_i\Sigma_j(\Sigma_k y_{hijk}^2 - T_{hij.}^2/r) \qquad (4.3.2)$$

The mean square

$$M_r = S_r / bwp(r-1)$$

is an unbiased estimate of σ_r^2 with bwp(r-1) degrees of freedom.

Averages for placements can be approached in same way. If $T_{hi.}$ is written for the total of the observations on the i^{th} wafer in the h^{th} batch, the sum of squares between placements (on the same wafer) is:

$$S_p = \Sigma_h\Sigma_i\Sigma_j T_{hij.}^2/r - \Sigma_h\Sigma_i T_{hi..}^2/pr$$

The mean square for placements is:

$$M_p = S_p / (bw(p-1))$$

and

$$E(M_p) = \sigma_r^2 + p\sigma_p^2$$

The sum of squares between wafers is

$$S_w = \Sigma_h\Sigma_i T_{hi..}^2 - \Sigma_h T_{h...}^2$$

and

$$E(M_w) = E(S_w / (b(w-1)) = \sigma_r^2 + r\sigma_p^2 + pr\sigma_w^2$$

Finally, the sum of squares between batches is:

$$S_b = \Sigma_h T_{h...}^2 - G^2 / N$$

where G is the grand total of the N observations.

Then

$E(M_b) = E(S_b / (b-1)) = \sigma_r^2 + r\sigma_p^2 + pr\sigma_w^2 + wpr\sigma_b^2$

These calculations are summarized in the traditional ANOVA table, shown below as Table 4.3.1.

Table 4.3.1 ANOVA for experiment B (nested)

Source	**DF**	**EMS**
batches	b-1	$\sigma_r^2 + r\sigma_p^2 + pr\sigma_w^2 + wpr\sigma_b$
wafers	b(w-1)	$\sigma_r^2 + r\sigma_p^2 + pr\sigma_w^2$
placements	bw(p-1)	$\sigma_r^2 + \sigma s_p^2$
repeat	bwp(r-1)	σ_r^2

When the values of M_b, M_w, M_p, and M_r are calculated from the data and equated to the expected values that are given in Table 4.3.1, there are four equations to be solved for the four components of variance.

The following comments on the calculations in Table 4.3.1 should be made. The calculations outlined above are straightforward because the design was balanced. There were exactly w wafers from each batch, exactly p placements for each wafer, and exactly r observations for each placement. When this balance is not present, the calculations are more complex. This topic will be discussed later in sections 4.5, 4.6, and 4.7.

It is possible that some of the estimates of the components will be negative. This fact is, on the surface, surprising since a variance cannot be negative. However, if variance is actually zero, one would not be surprised to find negative estimates. Since an unbiased method of estimation gives the expected answer (zero) *on the average*, it will sometimes give positive and at other times negative estimates.

The usual procedure when a negative estimate is obtained is to set that component of variance equal to zero, which is the same as dropping that term from the model.One can then delete the corresponding line from the ANOVA table and recalculate the estimates of the other variance components.

4.4 Missing Data in Nested Designs

Sections 4.5, 4.6, and 4.7 are devoted to calculating the expected mean squares in nested designs when there are missing data.

4.5 Nested Designs With Two Stages

Suppose that there are w wafers and that n_i observations are made on the i^{th} wafer. Let the variance between observations on the same wafer be σ_r^2 and let the variance between wafers be σ_w^2.

Let y_{ij} denote the j^{th} observation on the i^{th} wafer; the model for y_{ij} is:

$$y_{ij} = \mu + w_i + e_{ij}$$

where w_i is $N(0, \sigma_w^2)$ and e_{ij} is $N(0, \sigma_r^2)$; and w_i and e_{ij} are independent random variables. Let W_i denote the total for the i^{th} wafer and let $N = \Sigma_i n_i$; and N is the total number of observations.

Then:

$$E(y_{ij}^2) = \mu^2 + \sigma_r^2 + \sigma_w^2$$

and:

$$G = N\mu + \Sigma n_i w_i + \Sigma\Sigma e_{ij}$$

$$E(G^2)/N = N\mu^2 + \Sigma n_i^2 \sigma_w^2/N + \sigma_r^2$$

so that:

$$E(SStotal) = N(\mu^2 + \sigma_0^2 + \sigma_1^2) - N\mu^2 - \Sigma n_i^2 \sigma_1^2/N - \sigma_0^2$$

$$= (N-1)\sigma_r^2 + (N - \Sigma n_i^2/N)\sigma_w^2$$

Also:

$$W_i = n_i\mu + n_i w_i + \Sigma_j e_{ij}$$

and:

$$E(W_i^2) = n_i^2\mu^2 + n_i^2\sigma_w^2 + n_i\sigma_r^2$$

so that:

$$E(SSwafers) = \Sigma E(W_i^2/n_i) - E(G^2/N)$$

$$= N\mu^2 + N\sigma_w^2 + w\sigma_r^2 - N\mu^2 - \Sigma n_i^2\sigma_w^2/N - \sigma_r^2$$

$$= (w-1)\sigma_r^2 + (N - \Sigma n_i^2/N)\sigma_w^2.$$

These calculations are combined to obtain the analysis of variance shown in Table 4.5.1.

Table 4.5.1 Two-stage nested design with missing data

Source	**DF**	**SS**	**EMS**
Wafers	w-1	$\Sigma W_i^2/n_i - G^2/N$	$\sigma_r^2 + k_1\sigma_w^2$
Residual	N-w	$\Sigma\Sigma y_{ij}^2 - \Sigma W_i^2/n_i$	σ_r^2

where $k_1 = (N - \Sigma n_i^2/N)/(w-1)$. There is no difficulty about the usual F test for the hypothesis that $\sigma_w^2 = 0$.

Note: In the balanced case where there are w wafers and n observations per wafer N = nw and n_i = n for all i, so that:

$$k_1 = (nw - n^2w/nw)(w-1) = n$$

4.6 Three Stage Nested Design With Missing Data

The calculations with three stages are similar but more tedious and will not be treated in detail here.

Suppose these are b batches of wafers with m_i wafers in the i^{th} batch and that there are n_{ij} observations on the j^{th} wafer in the i^{th} batch. Let $W = \Sigma_i m_i$ denote the total number of wafers; let n_i

$= \Sigma_j n_{ij}$ denote the number of observations on the i^{th} batch, and let $N = \Sigma_i \Sigma_j n_{ij}$ denote the total number of observations.

The analysis of variance table is shown as Table 4.6.1.

Table 4.6.1 Three-stage nested design with missing data

Source	**DF**	**SS**	**EMS**
Batches	b-1	$\Sigma B_i^2/n_i - \Sigma\Sigma W_{j(i)}^2/n_{ij}$	$\sigma_r^2 + k_2\sigma_w^2 + k_3\sigma_b^2$
Wafers	W-b	$\Sigma\Sigma W_{j(i)}^2/n_{ij} - G^2/N$	$\sigma_r^2 + k_1\sigma_w^2$
Residual	N-W	by subtraction	σ_r^2

where

$k_1 = (N - \Sigma_i \Sigma_j (n_{ij}^2/n_i))/(W-b)$

$k_2 = (\Sigma_i \Sigma_j (n_{ij}^2/n_i) - \Sigma_i \Sigma_j n_{ij}^2/N) /(b-1)$

$k_3 = (N - \Sigma n_i^2/N)/(b-1)$

The F test for σ_b^2 now has problems. The ratio of the mean squares for batches to the mean square for wafers does not give a straightforward test because the expected values of the numerator and the denominator do not differ only in a multiple of σ_w^2.

4.7 Three-stage Calculations Procedure

The following is a suggested procedure for three-stage calculations

1. Perform a one way ANOVA for the b batches ignoring wafers and obtain SS (batches ignoring wafers); call this S_1.
2. Perform a one way ANOVA for the W wafers and get sum of squares for wafers ignoring batches and the residual sum of squares. Call these S_2 and S_3, respectively.
3. Then, in Table 4.6.1, SS(batches) = S_1, SS(wafers) = $S_2 - S_1$ and the residual sum of squares is S_3.

5 SETTING UP MODELS FOR FACTORIAL EXPERIMENTS AND CALCULATING EXPECTED SUMS OF SQUARES

This topic is illustrated with three examples:

1. A straightforward crossed classification.
2. A simple nested design.
3. A more complex design in which there are both crossed and nested factors.

5.1 Fixed and Random Factors

Recalling the definitions of Milliken and Johnson [5]:

A factor is *random* if its levels are a random sample of levels from a population of possible levels.

A factor is *fixed* if its levels are selected by a non-random process or if the levels consist of the entire population of possible levels.

Deciding whether a factor is fixed or random is usually easy, but not always so. It seems reasonable to regard the wafers in this experiment as a random sample of the infinite population of wafers that could be made under the specified conditions. Suppose, on the other hand, that nitric, sulphuric, and hydrochloric acids are compared. These three compose the whole list of acids under consideration and all have been selected; acids is a fixed factor.

It is customary to act as if the population the levels of a random factor are chosen from is infinite or, at least, very large. If lots contain only 24 wafers and a sample of six wafers are chosen at random, six out of the 24 wafers are selected. The investigator does not worry about sampling from a finite population. The reader who is concerned about this point should read Cornfield and Tukey [6]. They considered the population size to be N in their calculations; then they let N equal the actual number of levels for a fixed factor and let N tend to infinity for a random factor.

It was mentioned in Section 3 that operators may be classified either as a fixed or a random factor in different situations. If the experimental design called for two operators, *operators* might be regarded as a random factor so that the results might apply to any operator anywhere who might use such a machine. Are the two operators who were actually used in an experiment really chosen at random, and, if so, from what population? Were Arnie and Bill a random sample of all possible operators who have access to this measurement equipment? Were they merely the only two operators in our lab? Were they a random sample from a population of three (Charlie was out sick)?

Sites is usually taken to be a fixed factor. Experimenters usually choose specific sites, such as the center, and the traditional sets of four, two-five in an inner ring and six-nine in an outer ring, and measure the thickness of the film at the same place each time. With that assignment, it is reasonable to regard sites as a fixed factor. On the other hand, if the engineer made five observations at random places on the wafer, *sites* would be a random factor.

The importance of this distinction will be investigated later.

A model for a factorial experiment is called *fixed* if all the factors are fixed, *random* if all the factors are random, and *mixed* if the experiment contains both fixed and random factors. Note that the division into fixed, random, and mixed models does *not* depend on whether factors are crossed or nested (see Section 5.2).

5.2 Crossed and Nested Factors

Two factors, A and B, are crossed if the same levels of A appear with each level of B, and vice versa. B is said to be nested within A if the levels of B are different for each level of A.

The following are two classical examples of crossed and nested (hierarchic) experiments.

A typical experiment with three factors all crossed could involve temperature, pressure, and acid strength with three levels of temperature (100, 150, and 200 degrees); and four levels of pressure (375, 400, 425, and 450 psi) and two acids (HNO_3, HCl) with two observations made at each set of conditions or treatment combination. There are 3x4x2x2 = 48 total observations. All three factors are fixed. The main purpose of the experiment is to compare temperatures, pressures, and acids. Perhaps the experimenter wishes to fit curves of yield against temperature and to find which settings on the three dials produce the highest yield.

If, on the other hand, the four pressures were replaced by four barrels of crude oil, *barrels* would be a random factor and the model would be mixed. It would not make any difference in the various sums of squares and their degrees of freedom. It would make a difference in the expected values of some of the sums of squares.

The classic examples of hierarchical designs with several nested factors come from agriculture with sires, dams and siblings. An industrial example might have f formulations of gasoline, b batches made of each formulation, s samples taken from each batch, and n analyses (octane number determinations) made on each sampling. In this example, batches are nested in formulations, samples are nested in batches, and analyses are nested in samples. Formulations might be random or fixed depending on how they were chosen. It makes no difference to the sums of squares. The main interest in the batches and samples is to estimate their components of variance. The sum of squares within samples gives an estimate of the *repeatability* of the machinery that estimates the octane number. Note that, in this case, *repeatability* tells only about the variability in the procedure to estimate octane numbers. It tells nothing about the variability in the sampling procedures or in the refining system.

Section 5.3 begins a discussion of the crossed experiment, or the experiment in which all the factors are crossed. In the simplest case, all the factors are fixed factors. Section 3.3 contained an example of a crossed experiment with two factors, both random; that will lead to Section 5.4. Section 5.5.3 shall consider the crossed experiment in which some factors are fixed and some are random. Section 5.6 will return to hierarchic experiments.

5.3 The Crossed Experiment

To write the model for the crossed experiment with three factors, the factors are denoted by A, B, C with a, b, c levels respectively and there are n observations at each combination of the factors. Let y_{ijkm} denote the m^{th} observation at the i^{th} level of A, the j^{th} level of B, and the k^{th} level of C. The model has terms for all the main effects and all the interactions:

$$y_{ijkm} = \mu + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + (abc)_{ijk} + e_{m(ijk)}$$

a_i is the main effect of the i^{th} level of factor A; $(ab)_{ij}$ denotes the interaction between the i^{th} level and the j^{th} level of B; and $(abc)_{ijk}$ represents the three-factor interaction. The subscripts refer to the factors that appear in the term; and thus the A*C interaction has subscripts i and k. There is an interaction for every combination of two or three of the letters A, B, and C.

5.4 Rule for Degrees of Freedom

To find the number of degrees of freedom in a sum of squares one replaces each subscript by the number of levels in the corresponding factor minus one and then takes the product. Thus the sum of squares for A has (a-1) d.f, the AB interaction has (a-1)(b-1) d.f, and the ABC interaction has (a-1)(b-1)(c-1) d.f.

The total of the degrees of freedom is abcn-1. The number of d.f. for error is obtained by subtracting the d.f. for the other terms in the model from this total and getting, in this case, abc(n-1). Alternatively, it can be argued that there are n observations in each cell with n-1 d.f.; and multiplying by the number, abc, of cells gives abc(n-1) d.f.

5.5 Expectations of the Mean Squares

So far in this experiment, no distinction has been made between fixed and random factors. That distinction appears in the e.m.s. The fixed and random models are the extreme cases. The mixed models fall between the two extremes.

5.5.1 Fixed Models

If all the effects are fixed, the expected value for the mean square of any term consists of the error variance plus the contribution of that particular term. The e.m.s. in this example are shown in Table 5.5.1.1.

Table 5.5.1.1 A, B, C Fixed

Source	**DF**	**EMS**
A	a-1	σ^2 + Q[A]
B	b-1	σ^2 + Q[B]
C	c-1	σ^2 + Q[C]
AB	(a-1)(b-1)	σ^2 + Q[AB]
AC	(a-1)(c-1)	σ^2 + Q[AC]
BC	(b-1)(c-1)	σ^2 + Q[BC]
ABC	(a-1)(b-1)(c-1)	σ^2 + Q[ABC]
Error	abc(n-1)	σ^2

The expression Q[] stands for some quadratic expression in the corresponding terms in the model. For example:

$$Q[A] = bcn\Sigma a_i^2 / (a-1)$$

To test the hypothesis that all the a_i are zero the mean square M_A is compared to the error mean square, s^2. In this example, every term is tested against the mean square for error.

5.5.2 Random Models

If all the factors are random, the expected mean squares contain many more terms. The e.m.s. are shown in Table 5.5.2.1.

Table 5.5.2.1 All Factors Random

Source	**EMS**
A	$\sigma^2 + n\sigma^2_{ABC} + cn\sigma^2_{AB} + bn\sigma^2_{AC} + bcn\sigma^2_A$
B	$\sigma^2 + n\sigma^2_{ABC} + cn\sigma^2_{AB} + an\sigma^2_{BC} + acn\sigma^2_B$
C	$\sigma^2 + n\sigma^2_{ABC} + bn\sigma^2_{AC} + an\sigma^2_{BC} + abn\sigma^2_C$
AB	$\sigma^2 + n\sigma^2_{ABC} + cn\sigma^2_{AB}$
AC	$\sigma^2 + n\sigma^2_{ABC} + bn\sigma^2_{AC}$
BC	$\sigma^2 + n\sigma^2_{ABC} + an\sigma^2_{BC}$
ABC	$\sigma^2 + n\sigma^2_{ABC}$
Error	σ^2

All the components of variance for main effects and interactions are considered random and are expressed as variances. Thus $\sigma_{AB}{}^2$ is the expectation of $(ab)_{ij}{}^2$.

The terms in each row of the table are included according to the following rules:

- Each expected mean square contains as its first term σ^2.
- Each of the other terms in an expected mean square (component of variance) contains some letters as subscripts. The coefficient of any component is the product of all the letters that do not appear among its subscripts. With three factors there are four letters a, b, c, n. Hence the coefficient of σ^2_C is abn and the coefficient of σ^2_{AC} is bn.
- If every factor is random, the expectation of a mean square consists of σ^2 together with all the terms for main effects and interactions that contain *all* the letters which occur in the name of the particular mean square. For example $E(S_{BC})$ contains all the terms that have both B and C; and hence:

$$E(S_{BC}) = \sigma^2 + n\sigma^2_{ABC} + an\sigma^2_{BC}.$$

5.5.3 Mixed Models

If some factors are fixed and some are random the e.m.s. are obtained by adding these two rules to the three given for random models:

- If some of the factors are fixed, a distinction is made between fixed and random interactions and defined as follows. An interaction is considered to be fixed if *all* its factors are fixed; otherwise it is considered to be random.

  In the mixed model the e.m.s. for the random model is written down. Then any fixed interaction not the title of that row is struck out.

- In the terms for fixed effects, $bcn\sigma_A^2$ is replaced by Q[A] and so on.

If A as fixed and B and C were random, the e.m.s. would be unchanged from the random model except that $bcn\sigma_A^2$ would be replaced by Q[A]. However, if A were random and B and C were fixed, there would be one important change: the BC interaction is fixed and struck from the rows B and C. The e.m.s. are given in Table 5.5.3.1.

Table 5.5.3.1 A Random, B and C Fixed

Source	**EMS**
A	$\sigma^2 + n\sigma^2_{ABC} + cn\sigma^2_{AB} + bn\sigma^2_{AC} + bcn\sigma^2_A$
B	$\sigma^2 + n\sigma^2_{ABC} + cn\sigma^2_{AB} + Q[B]$
C	$\sigma^2 + n\sigma^2_{ABC} + bn\sigma^2_{AC} + Q[C]$
AB	$\sigma^2 + n\sigma^2_{ABC} + cn\sigma^2_{AB}$
AC	$\sigma^2 + n\sigma^2_{ABC} + bn\sigma^2_{AC}$
BC	$\sigma^2 + n\sigma^2_{ABC} + Q[BC]$
ABC	$\sigma^2 + n\sigma^2_{ABC}$
Error	σ^2

5.6 Hierarchic Models

In the example with formulations, batches, samples, and analyses, y_{ijkm} denotes the m^{th} analysis on the k^{th} sample from the j^{th} batch made with the i^{th} formulation. The appropriate model would be:

$$y_{m(ijk)} = \mu + f_i + b_{j(i)} + s_{k(ij)} + e_{m(ijk)}$$

The parentheses denote the nesting: $b_{j(i)}$ denotes that batches are nested in formulations (j nested in i), and $s_{k(ij)}$ denotes that samples are nested in both batches and formulations, while analyses are nested in all three previous steps.

In the crossed experiment model there are interactions for every combination of two or three (or more) letters. When there are nested factors there is an important restriction: no factor can interact with a factor within which it is nested. In this example the new rule means that there are no interactions, but more complex situations will be seen later.

The degrees of freedom are calculated as in the crossed case with the following crucial exception. Earlier the subscript i was replaced by (a-1) degrees of freedom; that rule is now extended. If a subscript is <u>not</u> in parentheses it is replaced by the number of levels minus one. If a subscript is in parentheses it is replaced by the number of levels. The degrees of freedom and the expected mean squares for the example are shown in Table 5.6.1.

In calculating the e.m.s., one uses the final rule of Section 5.5.2. The e.m.s. for formulations contains the repeatability, F, B(F), and S(BF) terms. The e.m.s. for batches contains the repeatability and B(F) and S(BF) terms. The e.m.s. for samples contains only the repeatability and the S term.

Table 5.6.1 Nested or Hierarchical Design

Source	**DF**	**EMS**
formulations	f-1	$\sigma_e^2 + n\sigma_{S(BF)}^2 + sn\sigma_{B(F)}^2 + bsn\sigma_F^2$
batches	f(b-1)	$\sigma_e^2 + n\sigma_{S(BF)}^2 + sn\sigma_{B(F)}^2$
samples	bf(s-1)	$\sigma_e^2 + n\sigma_{S(BF)}^2$
analyses	bsf(n-1)	σ_e^2

If the formulations are a fixed factor, $bsn\sigma_F^2$ is replaced with Q[F].

5.7 Calculating the EMS in a more Complex Example

In the statement of the first rule in Section 5.5.3, an interaction is defined to be fixed if *all* its factors are fixed. Then rule 5 is carried out. This version of the first rule of Section 5.5.3 is called the unrestricted rule. It has superseded the restricted rule advocated earlier by Bennett and Franklin [7], Scheffé [8], and John [9].

The difference between the two rules is that the restricted rule calls an interaction fixed if *any* of its factors are fixed. Thus the restricted rule is more liberal in striking out interactions from expected mean squares than the unrestricted rule. The unrestricted rule is used by SAS and is the default choice in Minitab.

The unrestricted rule will be used in the following example that involves both crossed and nested factors.

L lots are taken (random); w wafers are chosen from each lot (random), and c cycles of observations at s (fixed) sites on each wafer are made. For the moment, cycles are assumed to be random. Three steps are followed:

1. The model is written. This is the most important step. Computer programs can do the rest, but the model must be correct.

   In the model y_{ijkm} denotes the observation on the m^{th} cycle at the k^{th} site on the j^{th} wafer from the i^{th} lot. Then,

   $$y_{ijkm} = \mu + l_i + w_{j(i)} + s_k + c_m + (ls)_{ik} + (lc)_{im} + (ws)_{jk(i)}$$
   $$+ (wc)_{jm(i)} + (sc)_{km} + (lsc)_{ikm} + (wsc)_{jkm(i)} + e_{ijkm}$$

   Recalling the earlier rules, no factor interacts with another factor in which it is nested. Because there was only one observation on each site per cycle the WxSxC(L) term will be used for error.

2. The degrees of freedom are calculated as in Section 5.6. They are shown in Table 5.7.1.

   If the degrees of freedom do *not* add up to N-1 where N is the total observations the model is wrong and there is probably a term missing from the list of interactions.

Table 5.7.1 Degrees of Freedom

Lots	l_i	L-1
Wafers	$w_{j(i)}$	L(W-1)
Sites	s_k	S-1
Cycles	c_m	C-1
LxS	$(ls)_{ik}$	(L-1)(S-1)
LxC	$(lc)_{im}$	(L-1)(C-1)
WxS(L)	$(ws)_{jk(i)}$	L(W-1)(S-1)
WxC(L)	$(wc)_{jm(i)}$	L(W-1)(C-1)
SxC	$(sc)_{km}$	(S-1)(C-1)
LxSxC	$(lsc)_{ikm}$	(L-1)(S-1)(C-1)
WxSxC(L)	$(wsc)_{jkm(i)}$	L(W-1)(S-1)(C-1)

3. The e.m.s. are written, assuming that all the factors are random, following the rule that each line of the e.m.s. table contains all the terms that have as subscripts all the letters in the title of the line. See Table 5.7.1.

If some of the factors are fixed, the experimenter follows the rule of assuming interactions to be fixed if *all* their factors are fixed, and then striking them out of those rows in which they are not the title. For example, if sites and cycles are fixed, the SxC interaction is deleted from the e.m.s. for S and the e.m.s. for C; Q[S], Q[C], and Q[SC] are then put in the appropriate places.

Table 5.7.2 EMS with All Factors Random

Lots	$\sigma_e^2 + w\sigma_{LSC}^2 + s\sigma_{WC(L)}^2 + c\sigma_{WS(L)}^2 + ws\sigma_{LC}^2 + wc\sigma_{LS}^2 + sc\sigma_{W(L)}^2 + wsc\sigma_L^2$
Wafers	$\sigma_e^2 + s\sigma_{WC(L)}^2 + c\sigma_{WS(L)}^2 + sc\sigma_{W(L)}^2$
Sites	$\sigma_e^2 + w\sigma_{LSC}^2 + lw\sigma_{SC}^2 + c\sigma_{WS(L)}^2 + wc\sigma_{LS}^2 + lwc\sigma_S^2$
Cycles	$\sigma_e^2 + w\sigma_{LSC}^2 + lw\sigma_{SC}^2 + s\sigma_{WC(L)}^2 + ws\sigma_{LC}^2 + lws\sigma_C^2$
LxS	$\sigma_e^2 + w\sigma_{LSC}^2 + c\sigma_{WS(L)}^2 + wc\sigma_{LS}^2$
LxC	$\sigma_e^2 + w\sigma_{LSC}^2 + s\sigma_{WC(L)}^2 + ws\sigma_{LC}^2$
WxS(L)	$\sigma_e^2 + c\sigma_{WS(L)}^2$
WxC(L)	$\sigma_e^2 + s\sigma_{WC(L)}^2$
SxC	$\sigma_e^2 + w\sigma_{LSC}^2 + lw\sigma_{SC}^2$
LxSxC	$\sigma_e^2 + w\sigma_{LSC}^2$
WxSxC(L)	σ_e^2

5.8 Minitab Printouts

Tables 5.8.1 and 5.8.2 show the e.m.s. for this example both when all factors are random and when sites and cycles are fixed. There were three lots, four wafers per lot, five sites per wafer, and two cycles.

Table 5.8.1 All Factors Random

MINITAB Printout

```
Source         Expected Mean Square
1   L        (11) + 4(10) + 5(8) + 2(7) + 20(6) + 8(5) + 10(2) + 40(1)
2   W(L)     (11) + 5(8) + 2(7) + 10(2)
3   S        (11) + 4(10) + 12(9) + 2(7) + 8(5) + 24(3)
4   C        (11) + 4(10) + 12(9) + 5(8) + 20(6) + 60(4)
5   L*S      (11) + 4(10) + 2(7) + 8(5)
6   L*C      (11) + 4(10) + 5(8) + 20(6)
7   S*W(L)   (11) + 2(7)
8   C*W(L)   (11) + 5(8)
9   S*C      (11) + 4(10) + 12(9)
10 L*S*C     (11) + 4(10)
11 Error     (11)
```

Table 5.8.2 S and C Fixed, L and W Random

MINITAB Printout

```
Source         Expected Mean Square
             (using unrestricted model)

1   L        (11) + 4(10) + 5(8) + 2(7) + 20(6) + 8(5) + 10(2) + 40(1)
2   W(L)     (11) + 5(8) + 2(7) + 10(2)
3   S        (11) + 4(10) + 2(7) + 8(5) + Q[3,9]
4   C        (11) + 4(10) + 5(8) + 20(6) + Q[4,9]
5   L*S      (11) + 4(10) + 2(7) + 8(5)
6   L*C      (11) + 4(10) + 5(8) + 20(6)
7   S*W(L)   (11) + 2(7)
8   C*W(L)   (11) + 5(8)
9   S*C      (11) + 4(10) + Q[9]
10  L*S*C    (11) + 4(10)
11  Error    (11)
```

6 CONTRASTS AND SINGLE DEGREES OF FREEDOM

A primary concern in metrology studies is estimating the components of variation in the measurement system. To that end, factorial experiments are designed with fixed or random factors that may be either crossed or nested. Then estimates of the components of variance are calculated from the sums of squares in the analysis of variance. However, these sums of squares do not always represent only the random variation that can be properly associated with the measurement system. Two examples of this phenomenon will now be considered.

It is well known that the differences of oxide layer thickness on a wafer are not entirely due to random effects. There is also a radial component, possibly from centrifugal forces. One might expect the thickness *apart from the random error* to vary with the distance from the center of the wafer. This is called the bull's-eye effect. In the first example the contribution of the bull's-eye effect is eliminated from the sum of squares between sites, leaving a sum of squares that provides a better understanding of the random variation between sites.

In the second example, data collected over a period of consecutive cycles will be viewed and the contribution due to a linear time trend will be extracted from the sum of squares between cycles.

The technique used involves computing the sum of squares for a contrast. This is reviewed in Section 6.1.

6.1 Contrasts in the One Way Analysis of Variance

Suppose that the experimenter wishes to compare t treatments and that there are r observations on each treatment. The sum of squares between treatments is given by $\Sigma T_i^2/r - G^2 / rt$, where T_i is the total for the i^{th} treatment and G is the grand total for all the treatments; it has t-1 degrees of freedom.

A contrast between the treatments is a linear combination of the totals (this can be done with averages if preferred):

$$C = \Sigma a_i T_i$$

with $\Sigma a_i = 0$. If all the treatments are identical the expected value of C is zero. The variance of C is $V(C) = r\Sigma a_i^2 \sigma^2$.

An obvious example is the situation where there are five treatments of which the first is the control and the others are new formulations. The experimenter might well be interested in the contrast:

$$4T_1 - T_2 - T_3 - T_4 - T_5$$

Then the experimenter might wish to assess the importance of this contrast and to determine whether it explains the lion's share of the difference between treatments. To do this the sum of squares for that single contrast is calculated. It has one degree of freedom and is given by:

$$C^2 / (r\Sigma a_i^2)$$

Note that the denominator is the variance of C without the multiplier σ^2.

This single component sum of squares can be subtracted from the sum of squares between treatments calculated earlier leaving a sum of squares with t-2 degrees of freedom.

A second contrast:

$$C_2 = \Sigma b_i T_i$$

is said to be orthogonal (independent) to the first contrast if:

$$\Sigma a_i b_i = 0$$

An example would be $C_2 = T_2 + T_3 - T_4 - T_5$. The sum of squares for this orthogonal contrast can also be pulled from the total leaving a residual sum of squares that has t-3 degrees of freedom.

6.2 The Application to Sites on Wafers

Consider first the case in which there are five sites on each wafer, one at the center and the other four symmetrically placed in a ring at the same distance from the center as in Figure 6.2.1.

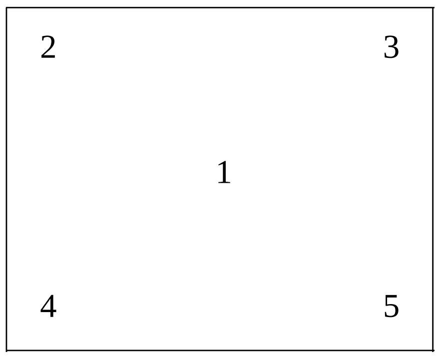

Figure 6.2.1 Diagram for Five Sites

The bull's-eye effect can be estimated by the first contrast used in Section 6.1 namely:

$$4T_1 - T_2 - T_3 - T_4 - T_5$$

If one argues that the bull's-eye effect is not random but inherent in the method of depositing the layer, the corresponding sum of squares for bull's-eye can be subtracted from the sum of squares for sites leaving a sum of squares with four degrees of freedom, and providing a valid estimate for the random variation in thickness over the wafer.

The difference between the left and right sides of the wafer can also be considered by the contrast:

$$T_2 - T_3 + T_4 - T_5$$

which is orthogonal to the previous contrast.

When there are nine sites, one at the center and four in each of two rings, as in Figure 6.2.2, there are more choices for contrasts.

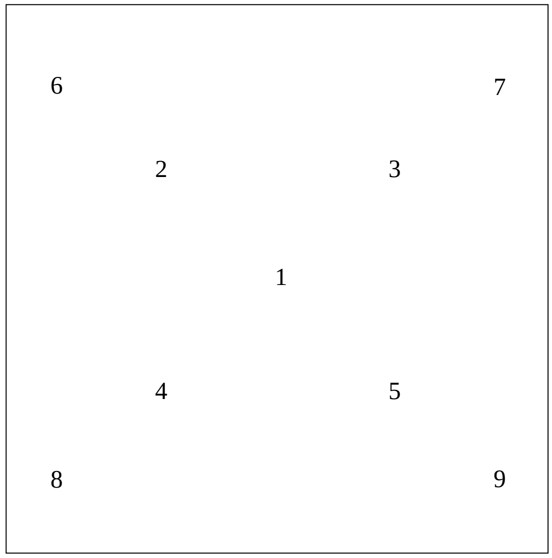

Figure 6.2.2 Diagram for Nine Sites

The bull's-eye effect can be evaluated by comparing the center point to the average of the other eight. This comparison is the contrast

$$8T_1 - T_2 - T_3 - T_4 - T_5 - T_6 - T_7 - T_8 - T_9$$

A second (orthogonal to the first one) contrast can then be taken that compares the average of the sites in the first ring to the average of the sites in the outer ring:

$$T_2 + T_3 + T_4 + T_5 - T_6 - T_7 - T_8 - T_9$$

The sums of squares for both contrasts can be extracted from the total sum of squares between sites.

Note this would not work if the following two contrasts are taken:

$$4T_1 - T_2 - T_3 - T_4 - T_5$$

and

$$4T_1 - T_6 - T_7 - T_8 - T_9.$$

The first compares the center to the first ring, and the second compares the center to the outer ring. However they are *not* orthogonal:

$\Sigma a_i b_i = 4x4 - 1x0 - 1x0 \ldots -1x0 = 16 \neq 0$. If they were used one could not subtract both sums of squares from the sum of squares between sites because of the lack of orthogonality.

6.3 An Example

This example is taken from an investigation of reflectivity in which 30 cycles of observations were made at nine sites on a wafer. The experiment as a two factor crossed factorial design with nine sites and 30 cycles (times). The analysis of variance table is shown as Table 6.3.1.

Table 6.3.1 Analysis of Variance Table

SOURCE	DF	SS	MS
times	29	0.1616	0.0056
sites	8	622.9849	77.8731
residual	232	1.0684	0.0046
total	269	624.2148	

6.4 Contrasts Between Sites

The sum of squares between sites is 622.98, which is almost all the total sum of squares. The totals of the 30 observations at each site are given in Table 6.4.1.

Table 6.4.1 Totals of 30 Observations on Reflectivity at Each of Nine Sites

Site	Total of 30 Observations
1	475.700
2	610.000
3	599.900
4	573.300
5	594.700
6	497.100
7	527.400
8	525.100
9	524.200

The value of the contrast

$$8T_1 - T_2 - T_3 - T_4 - T_5 - T_6 - T_7 - T_8 - T_9$$

is -646.1. Its sum of squares is $(646.1)^2/(72x30) = 193.26$.

The value of the second contrast

$$T_2 + T_3 + T_4 + T_5 - T_6 - T_7 - T_8 - T_9$$

is 304.1. Its sum of squares is $(304.1)^2/240 = 385.32$.

Together the two contrasts account for 578.58, which is 578.58 / 622.98 = 93% of the sum of squares between sites.

A sum of squares 578.58 with two degrees of freedom can be attributed to the radial changes in reflectivity. The remaining 44.40 with six degrees of freedom estimates the random variability in the reflectivity over sites.

6.5 Taking Out a Degree of Freedom For a Linear Time Trend

The same principle holds when the sum of squares between times is considered. There is evidence in the data of a linear time trend in the response, suggesting that either the reflectivity decreases over time or that the measurement system drifts as it warms up. Clearly this calls for further investigation. In the first case, it would be unfair to charge this source of variance between times to the measurement system. The sum of squares for linear trend can be removed from the sum of squares between times in a manner similar to that used in Section 6.4.

Figure 6.5.1 shows a plot of the totals of the observations made at the nine sites for each cycle against the cycle number which is a plot of reflectivity against time. It is clear from the illustration that the reflectivity decreases over time. Table 6.5.1 is part of the regression printout when a straight line fit is made for the reflectivity totals against time.

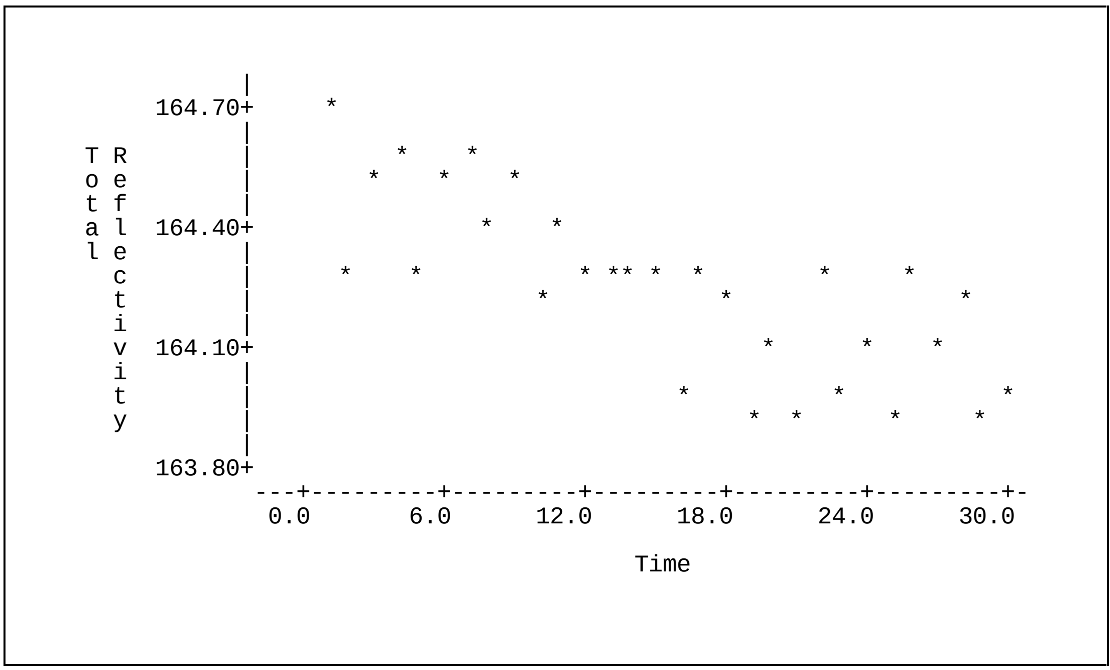

Figure 6.5.1 Total Reflectivity Over All Nine Sites vs. Time

Table 6.5.1 Regression of Reflectivity vs. Time

```
The fitted line is
          Total reflectivity = 164.551 - 0.020 (time)

Predictor        Coef         Stdev        t-ratio        p
Constant      164.551         0.054        3028.79    0.000
C22         -0.019622      0.003060          -6.41    0.000

s = 0.1451      R-sq = 59.5%

Analysis of Variance

SOURCE            DF         SS          MS        F        p
Regression         1    0.86533     0.86533    41.11    0.000
Error             28    0.58936     0.02105
Total             29    1.45468
```

The negative slope of the line is significantly different from zero. The sum of squares for regression is 0.86533.

The estimate of the slope is a contrast in the observations. The sum of squares for regression is the sum of squares for that contrast. It should be noted however, that in the fit totals of nine observations for the Y variable were used. In using these results in Table 6.5.1 the sum of squares for regression must be divided by nine. This means taking out for the single degree of freedom the quantity 0.086533 / 9 = 0.096.

The sum of squares for times can be separated into two components

SStime	29 DF	0.1616	m.s. 0.0056
linear trend	1 DF	0.86533/9 = 0.0961	m.s. 0.0961
remainder	28 DF	0.0655	m.s. 0.0023

6.6 The Amended ANOVA Table

Table 6.3.1 showed the original analysis of variance table. The extraction of contrast sums of squares with single degrees of freedom from both the sum of squares for sites and the sum of squares for times has changed that analysis of variance table. The new table is shown as Table 6.6.1.

Table 6.6.1 The Modified ANOVA Table

Source	DF	SS	MS
times	29	0.1616	0.0961
linear trend	1	0.0961	0.0023
remainder	28	0.0655	
sites	8	622.98	
1 vs others	1	385.32	385.32
inner vs. outer	1	193.26	193.26
rings	6	44.40	7.40
remainder			
error	232	1.0684	0.0046
total	269	624.2148	

6.7 Summary

In Section 6, the problem of orthogonal contrasts and their sums of squares has been considered and illustrated in two examples using the same data set both times. The primary purpose in both examples was to take a sum of squares that would normally be used for estimating random components of variance and to separate that sum of squares into two (or more) components. The first explains (and removes) the contribution of a source of nonrandomness.

In the first example, the source of nonrandomness was the bull's-eye effect, which is often present with such responses as the thickness of an oxide film on a wafer. When there were observations at only five sites of one at the center and four in a ring, only one contrast was taken. It compared the response at the center point to the average response on the ring. When there were nine sites in the center and two rings, two contrasts were considered: the center against the others and the inside ring against the outside ring.

A third orthogonal contrast could have been taken out corresponding to the right side of the wafer as opposed to the left, and a fourth corresponding to the *top* (sites 2, 3, 6, 7) as opposed to the *bottom* (sites 4, 5, 8, 9). In this example, the contrasts made a negligible contribution. One might be concerned that even after taking out the sums of squares for two orthogonal contrasts that represented the bull's-eye effect, the residual mean square was still 7.4. There is no obvious explanation of this. It is possible that the total for site six, 497.1, is unduly low. This might bear further investigation.

The experiment was a crossed experiment in which the measuring instrument made 30 cycles over all nine sites. The sum of squares between cycles should provide an estimate of the repeatability with 29 x 8 = 232 d.f. However, in this data set, there was clear evidence of a linear trend in the response. As the number of cycles increased, the reflectivity decreased. In the second example, a sum of squares with one d.f. that corresponded to the linear trend was taken out of the sum of squares between cycles. That left a sum of squares with 29 - 1 = 28 d.f. The mean square 0.0023 was compatible with the overall residual mean square, 0,0046.

This technique of partitioning a sum of squares into components that correspond to single degrees of freedom for orthogonal contrasts is commonly used in the analysis of variance. The reader is referred to the standard textbooks for further discussion. One of the most common applications occurs when one fits a polynomial model to a balanced experiment with a single factor. The sum of squares for that factor can be partitioned into contributions that correspond to orthogonal polynomials.

7 A QUICK METHOD OF DETERMINING THE REPRODUCIBILITY

Section 2 of this report described two basic experiments. In the first experiment, an operator took a wafer, loaded it in the machine and made n measurements at each site. The measurements are made in n cycles. The operator did *not* stop the measuring instrument between cycles. In the second experiment, the operator stopped the instrument after each cycle, took the wafer out, and loaded it again for the next cycle. The second experiment introduced an extra source of variability that was called the variance between placements and denoted by σ_p^2.

In this section, the theory of a method will be developed (proposed by Jack Reece of SEMATECH) for obtaining an easy and close approximation for the reproducibility in a crossed classification that is a version of the second experiment. This section concludes with some remarks about restricted and unrestricted models in the analysis of variance with mixed models.

7.1 The Crossed Classification Experiment

A wafer has s (fixed) sites. At t different times an instrument is set up to make n measurement cycles. This produces a data set with N = snt observations. The second experiment has n = 1.

The following model is used:

Let y_{ijk} denote the observation made on the i^{th} site on the k^{th} cycle at the j^{th} time.

$$y_{ijk} = \mu + s_i + t_j + (st)_{ij} + e_{k(ij)} \qquad (7.1.1)$$

The s_i are fixed effects with $\Sigma s_i = 0$; the t_j are random with mean zero and variance σ_j^2; the $(st)_{ij}$ are random with mean zero and variance σ_{st}^2; and the $e_{k(ij)}$ are random with mean zero and variance σ_r^2, where σ_r^2 is the repeatability. Note that σ_t^2 replaces the notation σ_p^2 that was used in Section 2.7.

Reece's procedure is to calculate individually for each site the sum of squares *between* the rt observations at that site. Each sum of squares has nt - 1 d.f. He then adds the sums of squares and divides the total by s(nt-1). The quotient is his estimate of total variability (reproducibility + repeatability).

This procedure is algebraically equivalent to the following:

- Making a one-way ANOVA on the sites, in which case his total will be the sum of squares *within* sites.

or

- Making the two-way crossed classification ANOVA for sites and times and pooling all the terms *except* sites, in which case his total will be the sum of the pooled terms.

The ANOVA table for the crossed classification is given in Table 7.1.1.

Table 7.1.1 Analysis of Variance

Source	**DF**	**EMS**
Sites	s-1	$\sigma_r^2 + n\sigma_{st}^2 + Q(s)$
Times	t-1	$\sigma_r^2 + n\sigma_{st}^2$
S x T	(s-1)(t-1)	$\sigma_r^2 + n\sigma_{st}^2$
Error	(n-1)st	σ_r^2

The sums of squares that are pooled have for their expectations:

Times $(t-1)(\sigma_r^2 + ns\sigma_t^2)$
S x T $(s-1)(t-1)(\sigma_r^2 + n\sigma_{st}^2)$
Error $(n-1)st\sigma_r^2$

Their total is:

$$s(nt-1)\sigma_r^2 + n(s-1)(t-1)\sigma_{st}^2 + ns(t-1)\sigma_t^2$$

It follows that the expectation of Reece's estimate of the total variability is:

$$E\left(\hat{\sigma}_j^2\right) = \sigma_r^2 + \left[n(s-1)(t-1)/s(nt-1)\right]\sigma_{st}^2 + \left[n(t-1)/(nt-1)\right]\sigma_t^2 \qquad (7.1.1)$$

If, for example, these are nine sites with 18 times and five cycles each time, so that

$$N = 810$$

$$E\left(\hat{\sigma}_j^2\right) = \sigma_t^2 + 0.85\sigma_{st}^2 + 0.955\sigma_e^2$$

results, which is close to

$$\sigma_t^2 + \sigma_{st}^2 + \sigma_r^2$$

The sum of the two components of variance

$$\sigma_t^2 + \sigma_{st}^2$$

is the reproducibility.

7.2 The Interaction Term

Can one go further and assume that $\sigma_{st}^2 = 0$? One good reason for this assumption is that it is difficult to justify such a term. This interaction term says that the *true* difference between sites varies with time. In a response such as thickness it is understood that there are differences between sites, for example, radially, but do these differences change with time? It could possibly occur if some areas on the wafer deteriorated more over time than others, but does that make practical sense? Dropping the interaction term leaves:

$$E\left(\hat{\sigma}_j^2\right) = \sigma_r^2 + \left[n(t-1)/(nt-1)\right]\sigma_t^2 \qquad (7.2.1)$$

Assuming that the interaction component is zero also avoids the conflict between the restricted and unrestricted models in the analysis of variance. This will be discussed in Section 7.3.

7.3 The Unrestricted Model

In constructing Table 7.1.1 the restricted model proposed by Scheffé [8] was assumed. It is given in most textbooks and used by many software packages. SAS, however, uses the unrestricted model. Minitab also uses the unrestricted model as the default but will use the restricted model on request. The difference lies in the values of the expected mean squares when there are random factors. For this two factor experiment the ANOVA with the unrestricted model when times are random and sites are fixed is shown in Table 7.3.1.

Table 7.3.1 Analysis of Variance Table (unrestricted model)

Source	**DF**	**EMS**
Sites	s-1	$\sigma_r^2 + n\sigma_{st}^2 + Q(s)$
Times	t-1	$\sigma_r^2 + n\sigma_{st}^2 + ns\sigma_t^2$
S x T	(s-1)(t-1)	$\sigma_r^2 + n\sigma_{st}^2$
Error	(n-1)st	σ_r^2

Calculating as before obtains:

$$E\left(\hat{\sigma}_j^{\,2}\right) = \sigma_r^{\,2} + \left[n(t-1)/(nt-1)\right]\sigma_{st}^{\,2} + \left[n(t-1)/(nt-1)\right]\sigma_t^{\,2} \qquad (7.3.1)$$

For the example with s= 9, t = 18, n = 4 this becomes:

$$E\left(\hat{\sigma}_j^{\,2}\right) = \sigma_r^{\,2} + 0.955\sigma_{st}^{\,2} + 0.955\sigma_t^{\,2}$$

This makes Reece's estimate look even better because it reduces the bias due to the interaction term.

7.4 An Example

In this data set, the wafer has five sites. On each of seven days four cycles were made and s = 5, t = 7, and n = 4.

Table 7.4.1 shows the mean squares, $[\Sigma_j\Sigma_k(y_{ijk} - \bar{y}_i)^2/27]$, for each site. The average of these mean squares is

$$\left(\hat{\sigma}_j^{\,2}\right) = 0.006033$$

If each mean square is multiplied by its d. f., nt-1 = 27, and the total is summed and divided by s(nt-1) = 135, the same estimate is obtained.

Table 7.4.1 Mean Square Deviations for Each Site

Site	MS
1	0.018878
2	0.004153
3	0.002153
4	0.002474
5	0.002509
average	**0.006033**

Table 7.4.2 shows the one-way ANOVA for the five sites with 28 observations per site. The sum of squares for error is 0.814493. The mean square for error is again 0.006033.

Table 7.4.2 One-way Analysis of Variance for Sites

Source	DF	SS	MS
sites	4	0.047696	0.011924
error	135	0.814493	0.006033

Table 7.4.3 shows the two-way ANOVA for sites and times. Note that, when the sums of squares for times, s x t, and error are added 0.814493 with 135 d.f. is again obtained.

Table 7.4.3 Two-way ANOVA

Source	DF	SS	MS
sites	4	0.047696	0.011924
times	6	0.299549	0.049925
s x t	24	0.119344	0.004973
error	105	0.395600	0.003768

The mean square for error is the estimate of the repeatability error, ($_r^2$) = 0.003768.

One can now turn to Table 7.1.1 and compute the estimates of the components of variance:

$$\left(\hat{\sigma}_t^2\right) = (0.049925 - 0.003768)/20 = 0.002308,$$

$$\left(\hat{\sigma}_{st}^2\right) = (0.004973 - 0.003768)/4 = 0.000301$$

($_{st}^2$) = (0.004973 - 0.003768) / 4 = 0.000301.

Finally, Equation 7.1.1 is confirmed by noting that for this data set it becomes, with s = 5, t = 7, n = 4,

$$\hat{\sigma}_j^2 = \hat{\sigma} + (32/45)\hat{\sigma}_{st}^2 + (8/9)\hat{\sigma}_t^2$$

that is,

$$0.006033 = 0.003768 + (0.711)(0.000301) + (0.889)(0.002308)$$
$$= 0.003768 + 0.000214 + 0.002052$$

One could conclude that $\sigma_{st}^2 = 0$ and drop it from the model. Then the interaction and error sums of squares could be pooled and the estimates of the other variance components could be recalculated as:

$$\hat{\sigma}_r^2 = (0.119344 + 0.395600)/129 = 0.003992$$
$$\hat{\sigma}_t^2 = (0.049925 - 0.003992)/20 = 0.006034$$

Then:

$$0.003992 + (0.889)(0.002297) = 0.006034$$

Similar results follow when the unrestricted model is used.

8 SUMMARY

Various scenarios have been discussed and models developed for them. As a summary, a generic design for gauge studies on wafers is suggested. This design can be modified along the lines of the previous sections for different situations. Among the assumptions made is that there is no *systematic* difference between operators in the sense that Charlie always tends to make measurements that are 10 Å higher than Dick's—because of the way he was trained. Given that assumption, there is no justification for involving more than one operator.

8.1 The Gauge Design

This time the engineer takes w (random) wafers, places each wafer in the machine on p occasions, and after each placement he makes a run. A run consists of r repeat measurements at each of s specific (fixed) sites. Altogether there are wp runs and N = wspr measurements.

Let y_{hijk} denote the k^{th} observation at the j^{th} site at the i^{th} placement of the h^{th} wafer;

$$1 \le h \le w,\ 1 \le i \le p,\ 1 \le j \le s,\ 1 \le k \le r$$

The model is longer than before:

$$y_{hijk} = \mu + w_h + p_{i(h)} + s_j + (ws)_{hj} + ps_{ij(h)} + e_{k(hij)} \qquad (8.1.1)$$

Placements are nested within wafers; p placements for each wafer gives w(p-1) d.f. for placements within wafers or P(W). The only fixed effect in the model is s. The other terms are random.

The data could be written in a rectangular array with s rows that correspond to the sites and wp columns that correspond to the runs and there would be r observations in each cell. That is a simple crossed classification, sites compared to runs. It would give the analysis of variance table outlined in Table 8.1.1.

Table 8.1.1 Crossed Classification Sums of Squares and d.f.

Runs	wp - 1
Sites	s - 1
Sites x runs	(wp-1)(s-1)
repeatability	wps(r-1)

When the model of Equation 8.1.1 is fitted to the data the sum of squares for runs in Table 8.1.1 is split into components for wafers with (w-1) d.f., and for placements (within wafers) with w(p-1) d.f. Similarly, the sum of squares for sites x runs becomes two component sums of squares for wafers x sites (W x S) and for placements x sites (within wafers). The terms for wafer x site and placement x site interactions reflect differences between measurements at various sites for different placements.

The numbers of degrees of freedom for the various terms are shown in Table 8.1.2. The expected values of the mean squares are shown in Table 8.1.3.

Table 8.1.2 Degrees of Freedom

Wafers	w-1
Placements (in W)	w(p-1)
Sites	s-1
W x S	(w-1)(s-1)
P x S (in W)	w(p-1)(s-1)
Repeatability	wsp(r-1)

Table 8.1.3 Expected Mean Squares

Wafers	$\sigma_r^2 + r\sigma_{ps(w)}^2 + pr\sigma_{ws}^2 + sr\sigma_{p(w)}^2 + psr\sigma_w^2$
Placements (in W)	$\sigma_r^2 + r\sigma_{ps(w)}^2 + sr\sigma_{p(w)}^2$
Sites	$\sigma_r^2 + r\sigma_{ps(w)}^2 + pr\sigma_{ws}^2 + Q[S]$
W x S	$\sigma_r^2 + r\sigma_{ps(w)}^2 + pr\sigma_{ws}^2$
P x S (in W)	$\sigma_r^2 + r\sigma_{ps(w)}^2$
Repeatability	σ_r^2

The reproducibility variance is estimated by:

$$\hat{\sigma}_r^2 = \hat{\sigma}_{p(w)}^2 + \hat{\sigma} h_{ps(w)}^2$$
$$= \left[M_{p(w)} + (s-1)M_{ps(w)} - sM_r\right] / sr$$

If the engineer wishes to test for site differences by taking out sums of squares with single degrees of freedom for the bull's-eye effect, etc., it can be done. That will make no difference to the estimates of repeatability and reproducibility.

8.2 Placements as Interaction

Suppose that the engineer elects to introduce days into the scenario. On each of p days he places each of the w wafers in the measuring instrument in turn and makes a run. How does the addition of days change the analysis? Mathematically this problem is similar to the basic experiment of Section 2, except that there are now days where previously there were operators.

There is still a crossed classification with wp runs and s sites. The difference between the analysis in this case and in Table 8.1.2 lies in the subdivision of the sum of squares between runs. This time the runs represent a crossed classification between days, and wafers. The subdivision is into three components: wafers, days, and wafers x days. The sum of squares for wafers is the same as it was in Table 8.1.1. The sum of squares for placements within wafers is divided into a sum of squares for days and a sum of squares for placements. The latter is algebraically identical with the sum of squares for interaction.

The reader acquainted with split-plot designs will note the similarity. One can regard the runs as whole plots, while the sites play the role of subplots. Then the sum of squares for placements is the same as the sum of squares for whole-plot error.

8.3 Should Wafers be Random or Fixed?

This paper treated wafers as a random effect. Does that matter? Perhaps the engineer carefully chose three wafers, one each at the high, middle, and low levels of the operating scale. Will this cause major changes in the analysis? The answer is *No*. The only difference will be that the term $prs\sigma_{ws}^2$ will disappear from the ems for wafers and for sites. It will not affect estimates of repeatability or reproducibility.

8.4 Caveats

Armed with a good computer program, the engineer can make various changes, such as electing to include terms for cycles and for interaction with cycles in the model, and including corresponding sums of squares in the analysis of variance table. The computer can also be used to calculate an approximate F test for differences between wafers when wafers are random because no exact F test exists. That is irrelevant to the fundamental question of metrology. It is not obvious that much good will come out of adding such complications.

What the engineer needs is a sound generic procedure for deciding whether a machine is satisfactory or unsatisfactory. The design proposed here should provide the engineer with such a procedure.

9 REFERENCES

[1] Barrentine, Larry B. (1991) *Concepts for R and R Studies*, ASQC Quality Press, Milwaukee. This carries the blessing of the American Society for Quality Control.

[2] Wheeler, Donald J. and Richard W. Lyday (1989) *Evaluating the Measurement Process*, 2nd edition, SPC Press, Inc., Knoxville.

[3] Wheeler, Donald J. (1992) *Problems with Gauge R and R Studies*. Proceedings of the 46th Annual Quality Congress.

[4] Montgomery, Douglas C. and George Runger, *Gage Capability and Designed Experiments*, unpublished technical publication 92-11.

[5] Milliken, G.A. and D.E.Johnson (1984), Analysis of Messy Data, vol 1, Designed Experiments, Van Nostrand Reinhold, New York.

[6] Cornfield, J. and J.W.Tukey (1956), "Average Values of Mean Squares in Factorials", Annals of Mathematical Statistics, 27, 907-949.

[7] Bennett, C.A. and N.J.Franklin (1956), Statistical Analysis in Chemistry and the Chemical Industry, John Wiley.

[8] Scheffé, H.(1959), The Analysis of Variance, John Wiley.

[9] John, Peter W.M.(1971), Statistical Design and Analysis of Experiments, Macmillan.

SEMATECH Technology Transfer
2706 Montopolis Drive
Austin, TX 78741

http://www.sematech.org