1

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file - Votre référence

Our file - Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

DETECTING ACTIVE EFFECTS IN UNREPLICATED FACTORIAL EXPERIMENTS

by

DEREK R. BINGHAM, B.Sc.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Mathematics and Statistics

Carleton University

Ottawa, Canada

December, 1994

© Copyright, 1994

Derek Bingham

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

THE AUTHOR HAS GRANTED AN
IRREVOCABLE NON-EXCLUSIVE
LICENCE ALLOWING THE NATIONAL
LIBRARY OF CANADA TO
REPRODUCE, LOAN, DISTRIBUTE OR
SELL COPIES OF HIS/HER THESIS BY
ANY MEANS AND IN ANY FORM OR
FORMAT, MAKING THIS THESIS
AVAILABLE TO INTERESTED
PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE
IRREVOCABLE ET NON EXCLUSIVE
PERMETTANT A LA BIBLIOTHEQUE
NATIONALE DU CANADA DE
REPRODUIRE, PRETER, DISTRIBUER
OU VENDRE DES COPIES DE SA
THESE DE QUELQUE MANIERE ET
SOUS QUELQUE FORME QUE CE SOIT
POUR METTRE DES EXEMPLAIRES DE
CETTE THESE A LA DISPOSITION DES
PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP
OF THE COPYRIGHT IN HIS/HER
THESIS. NEITHER THE THESIS NOR
SUBSTANTIAL EXTRACTS FROM IT
MAY BE PRINTED OR OTHERWISE
REPRODUCED WITHOUT HIS/HER
PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE
DU DROIT D'AUTEUR QUI PROTEGE
SA THESE. NI LA THESE NI DES
EXTRAITS SUBSTANTIELS DE CELLE-
CI NE DOIVENT ETRE IMPRIMES OU
AUTREMENT REPRODUITS SANS SON
AUTORISATION.

ISBN 0-315-98644-1

Canada

Name Derek Bingham

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

0463

U-M-I

SUBJECT TERM

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0777
Guidance and Counseling 0519
Health 0580
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language 0679
General 0289
Ancient 0290
Linguistics 0291
Modern 0401
Literature 0294
Classical 0295
Comparative 0297
Medieval 0298
Modern 0316
African 0591
American 0305
Asian 0352
Canadian (English) 0355
Canadian (French) 0593
English 0311
Germanic 0312
Latin American 0315
Middle Eastern 0313
Romance 0314
Slavic and East European

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion 0318
General 0321
Biblical Studies 0319
Clergy 0320
History of 0322
Philosophy of 0469
Theology 0323

SOCIAL SCIENCES

American Studies 0324
Anthropology 0326
Archaeology 0327
Cultural 0310
Physical 0272
Business Administration 0770
General 0454
Accounting 0338
Banking 0385
Management 0501
Marketing 0503
Canadian Studies 0505
Economics 0508
General 0509
Agricultural 0510
Commerce-Business 0511
Finance 0538
History 0366
Labor 0351
Theory 0578
Folklore 0366
Geography 0351
Gerontology 0578
History 0578
General

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science 0615
General 0616
International Law and Relations 0617
Public Administration 0814
Recreation 0452
Social Work 0626
Sociology 0627
General 0938
Criminology and Penology 0631
Demography 0628
Ethnic and Racial Studies 0629
Individual and Family Studies 0630
Industrial and Labor Relations 0700
Public and Social Welfare 0344
Social Structure and Development 0709
Theory and Methods 0999
Transportation 0453
Urban and Regional Planning 0453
Women's Studies

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture 0473
General 0285
Agronomy 0475
Animal Culture and Nutrition 0476
Animal Pathology 0359
Food Science and Technology 0478
Forestry and Wildlife 0479
Plant Culture 0480
Plant Pathology 0817
Plant Physiology 0777
Range Management 0746
Wood Technology 0306
Biology 0287
General 0308
Anatomy 0309
Biostatistics 0379
Botany 0329
Cell 0353
Ecology 0369
Entomology 0793
Genetics 0410
Limnology 0307
Microbiology 0317
Molecular 0416
Neuroscience 0433
Oceanography 0821
Physiology 0778
Radiation 0472
Zoology 0786
Biophysics 0760
General 0425
Biogeochemistry 0996
Geochemistry

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences 0566
General 0300
Audiology 0992
Chemotherapy 0567
Dentistry 0350
Education 0769
Hospital Management 0758
Human Development 0982
Immunology 0564
Medicine and Surgery 0347
Mental Health 0569
Nursing 0570
Nutrition 0380
Obstetrics and Gynecology 0354
Occupational Health and Therapy 0381
Ophthalmology 0571
Pathology 0419
Pharmacology 0572
Pharmacy 0382
Physical Therapy 0573
Public Health 0574
Radiology 0575
Recreation

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences 0485
Chemistry 0749
General 0486
Agricultural 0487
Analytical 0488
Biochemistry 0738
Inorganic 0490
Nuclear 0491
Organic 0494
Pharmaceutical 0495
Physical 0754
Polymer 0405
Radiation 0605
Mathematics 0986
Physics 0606
General 0608
Acoustics 0748
Astronomy and Astrophysics 0607
Atmospheric Science 0798
Atomic 0759
Electronics and Electricity 0609
Elementary Particles and High Energy 0752
Fluid and Plasma 0610
Molecular 0756
Nuclear 0611
Optics 0463
Radiation 0346
Solid State 0984
Statistics 0984
Applied Sciences 0346
Applied Mechanics 0984
Computer Science

Engineering 0537
General 0538
Aerospace 0539
Agricultural 0540
Automotive 0541
Biomedical 0542
Chemical 0543
Civil 0544
Electronics and Electrical 0348
Heat and Thermodynamics 0545
Hydraulic 0546
Industrial 0547
Marine 0794
Materials Science 0548
Mechanical 0743
Metallurgy 0551
Mining 0552
Nuclear 0549
Packaging 0765
Petroleum 0554
Sanitary and Municipal 0790
System Science 0428
Geotechnology 0796
Operations Research 0795
Plastics Technology 0994
Textile Technology

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

The undersigned recommend to the Faculty of Graduate Studies
and Research acceptance of the thesis

"DETECTING ACTIVE EFFECTS IN UNREPLICATED FACTORIAL EXPERIMENTS"

submitted by Derek R. Bingham, B.Sc.
in partial fulfillment of the requirements for
the degree of Master of Science

Thesis Supervisor

Chair, Department of Mathematics & Statistics

Carleton University

Abstract

Frequently in the manufacturing world, factorial experiments are used to investigate some aspect of a process or product. Due to resource constraints, the factorial experiment may be unreplicated. In these cases, experimenters have relied on half-normal probability plots (Daniel, 1959) to identify active effects, thus exposing the analysis to subjective biases. This thesis is concerned with automatic methods for identifying active effects in unreplicated 2^k factorial experiments. A simulation study is performed so that the performance of the automatic methods may be assessed.

Acknowledgements

I would like to thank Angela for her love and support, and my parents for always being there. I would also like to thank Randy Sitter for his guidance during my time at Carleton University.

Contents

Acceptance Sheet	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	ix
1 INTRODUCTION	1
1.1 Experimental Design	2
1.1.1 <i>Factorial Experiments</i>	3
2 METHODS FOR DETECTING ACTIVE EFFECTS	11
2.1 Common Methods for Detecting Active Contrasts in Factorial Experiments	12
2.1.1 <i>Half-normal Probability Plots</i>	12
2.1.2 <i>Using Pooled Inert Contrasts to Estimate the Error Variance</i> .	16
2.2 Automatic Methods for Detecting Active Effects	17
2.2.1 <i>Hamada and Wu's Method</i>	18
2.2.2 <i>Lenth's Method</i>	21
2.2.3 <i>Dong's Method</i>	28
2.2.4 <i>Benski's Method</i>	31

2.2.5	<i>Le and Zamar's Method</i>	35
3	Simulation Study	40
3.1	Simulation With Normal Yields (Uncalibrated)	42
3.1.1	<i>No Active Effects</i>	42
3.1.2	<i>One Active Effect</i>	45
3.1.3	<i>Two Active Effects</i>	47
3.1.4	<i>Three Active Effects</i>	48
3.1.5	<i>Discussion</i>	50
3.2	Simulation With Normal Yields (Calibrated)	51
3.2.1	<i>Calibrating the Methods</i>	52
3.2.2	<i>Simulation Study for Calibrated Methods</i>	53
3.2.3	<i>Simulation Results</i>	54
3.3	Simulation Study With Non-Normal Yields	60
3.3.1	<i>Simulation Study</i>	60
3.3.2	<i>Simulation Results</i>	61
4	A Permutation Test to Detect Active Effects	64
4.1	A Permutation Test	65
4.2	Simulation Study of the Permutation Method	68
4.2.1	<i>Simulation Results</i>	70
4.3	Revisiting Dong's Method	75
5	Conclusions	79
	Tables	83
	Figures	83
	Appendix	100

List of Tables

1	0.05 Significance Level Critical Values for Le and Zamar's Method . .	84
2	Simulation Results With No Active Contrasts	85
3	Simulation Results With One Active Contrast	86
4	Simulation Results With Two Active Contrasts	87
5	Simulation Results With Three Active Contrasts	89
6	Simulation Results For Calibrated Methods	92
7	Simulation Results With Non-Normal Yields	94
8	Simulation Results of Permutation Methods	97
9	Simulation Results of Step-up Procedure for Dong's Method	98

List of Figures

1	Half-normal probability plot of data in example 2.1	99
---	---	----

Chapter 1

INTRODUCTION

Competition in today's marketplaces has caused many companies to consider factorial experiments as a key to improvement of process or product quality. Deciding which factors affect the process or product can be an important step in achieving a competitive advantage. Frequently, due to resource or time constraints, the factorial experiments are unreplicated. In these cases, standard analysis of variance techniques may not be used to identify important factors. Daniel (1959) proposed a graphical method of analysis based on the quantiles of the half-normal distribution. Although widely used, this method relies on the experimenter's judgement and is therefore prone to experimenter bias. Thus, it is beneficial to have an automatic method, free of subjective biases, to determine which factors affect the process or product under investigation. This thesis is concerned with automatic methods for identifying the important factors in unreplicated factorial experiments.

This thesis consists of five chapters, this introductory chapter being the first. The remainder of this chapter will introduce the reader to experimental design and, in particular, to unreplicated factorial experiments. Chapter 2 will present the standard methods for analysis of factorial experiments and some recently proposed automatic methods for determining the important factors. The third chapter contains a simulation study of the new methods so that their relative abilities may be

examined.

Chapter 4 introduces a new automatic test (Loughin and Nobel, 1994) based on the permutation distribution of the yields of the factorial experiment. A small simulation study is performed so that the performance of this method may be compared to the simulation results in Chapter 3. The reader should note that this method was not included in the discussion in Chapter 2 and the simulation study in chapter 2. This is due to time constraints. The final chapter is a brief conclusion and summary.

1.1 Experimental Design

Often, due to the need for constant improvement in today's competitive marketplace, designed experiments are performed to discover something about a process. Designed experiments represent a series of tests in which changes are made so that reasons for variation in the process may be observed.

In the manufacturing world, the process may be the combination of machines and raw materials that produce some end product. This product may, in turn, have some measure placed on it (e.g., yield of the process, hardness of material, etc.) that indicates the state of the process.

Example 1.1: (Box, Hunter and Hunter, 1978)

An experimenter wishes to test the yield of a chemical process. Three factors are considered to affect the process - temperature, concentration and catalyst type.

The experimenter decides to test the process at two different temperatures, concentrations and catalysts.

In Example 1.1, it was the experimenter's goal to investigate the process by varying the three factors that may affect the process over various settings (or *levels*).

The measure of interest is the yield of the chemical process. Given the results of Example 1.1, the experimenter may wish to suggest new settings for the factors (*factor level settings*) that will result in an improved yield for the process.

Frequently, the objective of a designed experiment is to:

1. Determine which factors are most influential on the measure of interest,
2. Determine the optimal settings for the influential factors,
3. Determine the settings of the influential factors, so that the variability of the measure of interest is small (Montgomery, 1991).

The acquisition of such information may allow a manufacturer to improve its process yield or product quality, and thereby gain a competitive advantage in the marketplace.

1.1.1 *Factorial Experiments*

Before an investigation of a process is conducted, the experimenter knows that several factors may affect the process. Consequently, the experimenter is interested in the estimate of the impact of each of the factors in the experiment.

In the experiment in Example 1.1, information on the three factors can be found by varying the factor levels over all possible combinations of factor levels. That is, estimate the effect of each factor at all possible combinations of the other factors. Such an experiment is called a *factorial experiment*. The contribution of a factor of a factorial experiment is called the *main effect*. When a factorial experiment is performed, observations are collected at all possible factor level combinations and when one observation is observed at each of the possible factor level settings, the collection of these observations is called a *replicate* of the experiment.

Definition 1.1: *The estimated main effect of a factor, for an experiment with factors at two levels, is defined as the difference in the average response produced by the change in the level of the factor.*

The estimated main effect of a factorial experiment, calculated in this manner, can be thought of as the difference between the average response at the first level of a factor and the average response at the second level of the factor. When a factor has no effect on the response of the experiment, we expect this difference to be near zero, and the factor is said to be *inert*. Otherwise, the factor is said to be *active* (or *significant*).

Example 1.2:

The main effects of the three factors in Example 1.1 are to be estimated by varying the levels of each factor. The following data was observed for one replicate of the factorial experiment,

i	Temperature (T)	Concentration (C)	Catalyst (K)	Yield (y_i)
1	160	20	A	60
2	180	20	A	72
3	160	40	A	54
4	180	40	A	68
5	160	20	B	52
6	180	20	B	83
7	160	40	B	45
8	180	40	B	80

Using Definition 1.1, the main effect of the three factors are estimated as follows,

$$\text{Effect of T: } \frac{y_2+y_4+y_6+y_8}{4} - \frac{y_1+y_3+y_5+y_7}{4} = \frac{72+68+83+80}{4} - \frac{60+54+52+45}{4} = 23,$$

$$\text{Effect of C: } \frac{y_3+y_4+y_7+y_8}{4} - \frac{y_1+y_2+y_5+y_6}{4} = \frac{54+68+45+80}{4} - \frac{60+72+52+83}{4} = -5,$$

$$\text{Effect of K: } \frac{y_5+y_6+y_7+y_8}{4} - \frac{y_1+y_2+y_3+y_4}{4} = \frac{52+83+45+80}{4} - \frac{60+72+54+68}{4} = 1.5.$$

In many industrial experiments, practitioners may be interested in the study of two or more factors each at two levels. Often, factorial designs are employed in order to determine which factors are active and which factors are inert. For the remainder of this thesis, we will concentrate on methods of analyzing factorial experiments with factors at two levels¹.

Definition 1.2: *A factorial experiment with k factors, all with two possible factor level settings, is called a 2^k factorial experiment.*

It should be noted that for factorial experiments with all factors at two levels, rather than estimate the main effects individually, matrix algebra can be employed to compute the effect estimate of all factors at once. By coding the levels of each factor as a *High* and *Low* level, and representing these as coded values, $+1$ and -1 , the factor level settings for a run of a factorial experiment can be represented as ± 1 's. Consider a matrix X where each row is the coded values for the factor level setting of a run of a factorial experiment. Then the main effects may be estimated as follows,

$$(\text{effect } 1, \text{effect } 2, \dots, \text{effect } n) = \frac{1}{2^{k-1}} \mathbf{y}' \mathbf{X}$$

where,

$$\mathbf{y} = (y_1, y_2, \dots, y_{2^k})'.$$

¹In general, if an experiment has k factors and l_i factor levels for the i^{th} factor ($i = 1, 2, \dots, k$), then the experiment is said to be a $l_1 \times l_2 \times \dots \times l_k$ factorial design.

Example 1.3:

Consider the factorial experiment of Example 1.2. Rather than calculate the estimated main effects of the factorial experiment as shown in that example, matrix algebra may be used. The factor level settings of each run of the experiment are shown below:

Temperature (T)	Concentration (C)	Catalyst (K)
160	20	A
180	20	A
160	40	A
180	40	A
160	20	B
180	20	B
160	40	B
180	40	B

See that the levels of both concentration and catalyst may be referred to (arbitrarily) as High and Low, and that these levels may be coded as +1 and -1 respectively,

Temperature (T)	Concentration (C)	Catalyst (K)
-1	-1	-1
+1	-1	-1
-1	+1	-1
+1	+1	-1
-1	-1	+1
+1	-1	+1
-1	+1	+1
+1	+1	+1

The main effects can then be estimated as, $(T, C, K)' = \frac{1}{2^{3-1}} y'X$, where

$$X = \begin{pmatrix} -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ +1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & -1 & +1 \\ -1 & +1 & +1 \\ +1 & +1 & +1 \end{pmatrix},$$

is obtained from the above table.

By performing the factorial experiment, the experimenter is able to estimate the effect of each factor on the yield of the experiment. An added bonus of a factorial design is the ability to estimate the degree of interaction between the levels of the factors. The impact of the interaction of factors on the yield of a factorial experiment is called the *interaction effect*.

Definition 1.3: *In an experiment with factors at two levels, the estimate of the interaction effect is defined as the difference between the average effect of a factor at the high level of another factor(s) and the average effect of the factor at the low level of the other factor(s).*

The interaction effect measures the failure of the main effect of a factor to remain constant over the levels of another factor. Consider the following example.

Example 1.4: Consider Example 1.3. The interaction effect of the factors T and C may be estimated as follows:

By multiplying the coded levels of T and C for each factor level combination, a vector, X_{TC} , of +1's and -1's can be constructed. The estimate of the interaction effect may also be calculated by multiplying X_{TC} by y' and dividing by the number of contrasts at each level. For our example, we get

$$X_{TC} = \begin{pmatrix} -1 & \cdot & -1 \\ +1 & \cdot & -1 \\ -1 & \cdot & +1 \\ +1 & \cdot & +1 \\ -1 & \cdot & -1 \\ +1 & \cdot & -1 \\ -1 & \cdot & +1 \\ +1 & \cdot & +1 \end{pmatrix} = \begin{pmatrix} +1 \\ -1 \\ -1 \\ +1 \\ +1 \\ -1 \\ -1 \\ +1 \end{pmatrix}.$$

Then the estimated combined effect of T & C = $\frac{1}{4}y'X_{TC}$.

The analysis of a factorial experiment, such as the one in Example 1.2, may proceed by the use of *analysis of variance* techniques if replicates of the experiment are taken. That is, using the replicates of the factorial experiment, the observed sum of squares may be used to test whether an effect or interaction is inert. Evidence that a main effect or interaction is non-zero implies that the setting(s) of the factor(s) will affect the yield of the experiment.

If there exists only enough resources (e.g., budget, raw materials, etc.) for one replicate of the factorial experiment, then there is no estimate of the error variance and analysis of variance techniques cannot be used. Therefore, the experimenter is unable to test the hypothesis that an effect or contrast is zero using this technique. The remainder of this thesis will focus on methods of analyzing unreplicated 2^k factorial experiments. Toward this goal, we first introduce some general notation.

It has been noted that both the main effects and interaction effects can be estimated using matrix algebra. In order to generalize this notion, some notation

is required. Consider an unreplicated factorial experiment with k factors. Let the vector of responses for the 2^k unreplicated factorial design be,

$$\mathbf{y} = (y_1, y_2, \dots, y_{n+1})'$$

where $n = 2^k - 1$, and the matrix of coded levels representing the 2^k factor level settings be X_e . By multiplying the coded levels of pairs of factors, a column vector for each two factor interaction (or $2fi$'s) can be created. Augmenting all such columns gives the $2fi$ matrix X_{2fi} . Similarly, by multiplying the coded levels of groups of three factors the matrix X_{3fi} can be found. We can continue in this manner until matrix X_{nfi} is gained.

Augmenting these matrices with X_e , the *design matrix* X may be formed,

$$X = [X_e \mid X_{2fi} \mid X_{3fi} \mid \dots \mid x_{kfi}].$$

Let the vector,

$$\mathbf{c} = (c_1, c_2, \dots, c_n)$$

be the vector of main effects and interaction effects, and

$$\hat{\mathbf{c}} = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)$$

be the corresponding estimates based on $\mathbf{y}$. Then,

$$\hat{\mathbf{c}} = \frac{1}{2^{k-1}} \mathbf{y}' X.$$

At this point, it is the practitioner's goal to determine which of the main effects or interactions effects are active. That is, to decide which of the $c_i \neq 0$.

Often in the literature discussing factorial designs there is some inconsistency in the use of terms referring to effects and estimated effects. To avoid any confusion, for the remainder of the paper the main effects and interactions effects will be referred to as *effects*, and their estimates as *contrasts* (or simply *estimated effects*).

Therefore, the practitioner aims to identify active effects based on the observed contrasts. For the remainder of this thesis, the term *active contrast* will be used to indicate a contrast which estimates an active effect. Similarly, an *inert contrast* is a contrast that estimates an inert effect.

Chapter 2

METHODS FOR DETECTING ACTIVE EFFECTS

Factorial experiments are frequently used to identify the factors that effect a product or process. Due to cost constraints or limited experimental materials, it may be impossible to repeat the experiment numerous times. In these cases unreplicated factorial experiments are performed. Based on the results of the experiment, recommendations may be made to improve the product or process.

If there exists a *good* estimate of the error variance, the usual analysis of variance techniques may be used to determine whether a contrast is inert or active. However, often due to cost concerns or amount of experimental material, the experiment is unreplicated and no obvious estimate of the error variance exists. This chapter will present the common methods for identifying active effects as well as some recently proposed automatic methods for detecting active effects in unreplicated factorial experiments.

2.1 Common Methods for Detecting Active Contrasts in Factorial Experiments

2.1.1 Half-normal Probability Plots

Daniel (1959) proposed a graphical method to identify active contrasts in 2^k factorial experiments using the quantiles of the standard normal distribution. This method is still frequently used, and to some degree is the standard to which all other methods are compared. This section will provide an introduction to the half-normal probability plot. Consider a random sample, $(x_1, x_2, \dots, x_n)$ from the standard normal distribution. Let $(x_1^*, x_2^*, \dots, x_n^*)$ be the absolute values of the random sample and $(x_{(1)}^*, x_{(2)}^*, \dots, x_{(n)}^*)$ be the corresponding order statistics. The sample cumulative distribution function (cdf) can be written as

$$F_n(x_{(i)}^*) = \frac{i}{n}, \quad i = 1, 2, \dots, n,$$

or equivalently,

$$P_i = \frac{i}{n}, \quad i = 1, 2, \dots, n.$$

Written in this manner, we see that $x_{(1)}^*$ can be taken to represent the first $100(\frac{1}{n})\%$ of the cdf. Similarly, $x_{(2)}^*$ can be taken to represent the second $100(\frac{1}{n})\%$ of the cdf, and so on. Therefore,

$$P'_i = \frac{(i - 1/2)}{n}, \quad i = 1, 2, \dots, n$$

represent the midpoints of the intervals $[0, P_1], [P_1, P_2], \dots, [P_{n-1}, P_n]$.

Consider a graph where $(x_{(1)}^*, x_{(2)}^*, \dots, x_{(n)}^*)$ define the abscissa, and the ordinate is labeled with from 0 to 1. If the P'_i 's are plotted on this graph, then these values approximately trace out the sigmoid curve. However, if these points are plotted on half-normal probability paper (a grid where the spacing of the ordinate represent the spacing of the quantiles of the standard half-normal distribution), then the points $P'_1, P'_2, \dots, P'_n$ will fall on a straight line. This is expected since we are, in fact, plotting

the quantiles of the half-normal distribution versus the quantiles of a sample from the standard half-normal distribution. This type of graph is called a *half-normal plot*.

In fact, if we let $Z^* = (z_{(1)}^*, z_{(2)}^*, \dots, z_{(n)}^*)$ be the ordered absolute values of an iid sample drawn from a normal distribution with mean, 0, and variance σ^2 , then we expect the corresponding P_i' 's also to lie on a straight line since the standard Normal and the normal variates differ only by a scale factor, σ^2 .

If, however, the sample contains a value, z_k^* , from some other distribution, then we would not expect the corresponding P_k' to lie on the straight line. This would be observed because the location of P_k' along the abscissa, relative to the other P_i' 's, would not represent a quantile of the half-normal distribution. It is this property that Daniel employs to identify active contrasts in factorial experiments.

Using Half-normal Probability Plots to Identify Active Contrasts in Factorial Experiments

Consider an unreplicated factorial experiment with n effects, $c_1, c_2, \dots, c_n$. Let the contrasts based on the unreplicated factorial experiment be $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$. Consider the following theorem.

Theorem 2.1: Let $(y_1, y_2, \dots, y_{n+1})$ be yields of a factorial experiment, and $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$ be the corresponding contrasts. Then the $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$ are iid normally distributed if $(y_1, y_2, \dots, y_{n+1})$ are also iid normally distributed.

Proof: Suppose $(y_1, y_2, \dots, y_{n+1})$ are iid normally distributed with mean μ and variance τ^2 . Then the i^{th} contrast is

$$\hat{c}_i = \frac{1}{2^n} y' X_i$$

where X_i is the i^{th} column of the design matrix. See that $\hat{c}_i$ is a linear combination of iid normal variates. Therefore the $\hat{c}_i$'s are normally distributed with mean 0 and variance $[(n+1)\sigma^2]/2^{2n}$.

Consider any two of the contrasts, $\hat{c}_i$ and $\hat{c}_l$. Then

$$\text{cov}(\hat{c}_i, \hat{c}_l) = \sum_{j=1}^{n+1} x_{ji}x_{jl}\sigma^2.$$

By the orthogonality of the columns of the design matrix, $\sum_{j=1}^{n+1} x_{ji}x_{jl} = 0$. Therefore,

$$\text{cov}(\hat{c}_i, \hat{c}_l) = 0.$$

Since the contrasts are the linear combination of iid normal variates and have a covariance equal to zero, then the contrasts are also iid (by Lemma 5.4.2 of Casella and Berger, 1990). $\square$

Therefore, if we assume that the yields are iid normally distributed, then the contrasts of the factorial experiments are also iid normal.

If we assume that the contrasts are iid normally distributed (say, $c_i \sim N(0, \sigma^2)$ $\forall i = 1, 2, \dots, n$), we expect the sample cdf of the absolute contrasts to mirror the quantiles of the half-normal distribution. Therefore, if we construct a half-normal plot based on the absolute $\hat{c}_i$'s, then the corresponding P_i 's should all lie on a straight line. Points that do not exactly lie on a straight line may be attributed to noise within the sample.

If there is one active effect, c_k , then the remaining $(n-1)$ contrasts should correspond to P_i 's that lie on a straight line. We do not expect P_k to lie on the straight line of the half-normal plot since $\hat{c}_k$ is an active effect and therefore its contrast, $\hat{c}_k$, comes from some distribution other than $N(0, \sigma^2)$. Similarly, if there are two active effects, c_k and c_l , then P_k and P_l should not lie on the straight line of the normal probability plot of P_i 's. Using this procedure, Daniel suggested

constructing half-normal probability plots from the contrasts and those contrasts that do not lie on (or approximately on) a straight line are identified as active.

If a large proportion of the effects are active, then little can be learned from the half-normal probability plot. That is, in this case, we cannot expect the P_i 's to fall on a straight line since they do not mirror the quantiles of the half-normal distribution, and therefore, no conclusion can be made. However, it is often the case in practice that relative to the number of effects, n , the number of active effects is small. This property is referred to as *effect sparsity*. If effect sparsity is anticipated, half-normal probability plots offer a valuable tool for identifying active effects.

Some care should be taken when using half-normal probability plots. Until recently, there has been no automatic method to distinguish between the expected noise of the normal variates and contrasts estimating active effects that moderately deviate from the straight line. That is, one may find it very difficult to distinguish between an active contrast of relatively moderate magnitude and noise from the normal distribution, thus exposing the conclusions to subjective biases.

It should be noted that, rather than plot the P_i 's vs. the absolute contrasts on half-normal probability paper, frequently what is done is that the contrasts are plotted against the quantiles of the standard half-normal on half-normal probability paper. For the same reasons as discussed above, we expect the inert contrasts to lie on a straight line and the active contrasts to deviate from the straight line.

Example 2.1: (Box, Hunter and Hunter, 1981) A 2^4 factorial experiment was run and the following data observed.

Factor 1	Factor 2	Factor 3	Factor 4	Yield
-1	-1	-1	-1	71
+1	-1	-1	-1	61
-1	+1	-1	-1	90
+1	+1	-1	-1	82
-1	-1	+1	-1	68
+1	-1	+1	-1	61
-1	+1	+1	-1	87
+1	+1	+1	-1	80
-1	-1	-1	+1	61
+1	-1	-1	+1	50
-1	+1	-1	+1	89
+1	+1	-1	+1	83
-1	-1	+1	+1	59
+1	-1	+1	+1	51
-1	+1	+1	+1	85
+1	+1	+1	+1	78

The contrasts based on the above data are listed below.

i	1	2	3	4	5	6	7	8
Contrast	-8.00	24.00	-2.25	-5.50	1.00	0.75	0.00	-1.25

i	9	10	11	12	13	14	15
Contrast	4.50	-0.25	-0.75	0.50	-0.25	-0.75	-0.25

The half-normal probability plot of the contrasts are seen in Figure 1. See that while the two largest absolute contrasts are clearly active, it is difficult to assess the significance of the next largest absolute contrast.

2.1.2 *Using Pooled Inert Contrasts to Estimate the Error Variance*

If a good estimate of the error variance of the inert contrasts were available, then analysis of variance techniques could be used to test the significance of the contrasts.

As the experiment is unreplicated, no such estimate typically exists.

Suppose that $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$ are contrasts from a factorial experiment with no active effects. Then,

$$E(\hat{c}_i) = 0,$$

and,

$$\text{Var}(\hat{c}_i) = E[(\hat{c}_i - E(\hat{c}_i))^2] = E(\hat{c}_i^2)$$

for all $i = 1, 2, \dots, n$. Therefore, in this case where all contrasts are inert, an unbiased estimate of the error variance of the contrasts is

$$\widehat{\text{Var}} = \frac{\sum_{i=1}^n \hat{c}_i^2}{n-1}.$$

Based on this notion, it has been suggested that the error variance be estimated by successive pooling of supposedly inert contrasts (Quinlan, 1985). However, the difficulty lies in identifying the inert contrasts from which to calculate the error variance. Moreover, it turns out that this procedure will underestimate the variance, and thus overestimate the t -statistic used in analysis of variance techniques. This in turn will lead to falsely identified active contrasts (Box and Meyer, 1986), making this procedure unviable.

2.2 Automatic Methods for Detecting Active Effects

Identification of active effects in unreplicated factorial experiments may be difficult. Experimenters often rely on half-normal probability plots or pooled estimates for the error variance of the inert contrasts to help detect the presence of active effects. Unfortunately, these methods are frequently subjective in nature and prone to experimenter bias.

Recently, a number of automatic methods have been proposed to identify active effects in unreplicated factorial experiments. The aim of these methods is to eliminate the subjective biases of the half-normal probability plots. The following sections describe these automatic methods.

2.2.1 *Hamada and Wu's Method*

Consider an unreplicated factorial experiment with n effects, $c_1, c_2, \dots, c_n$, and contrasts $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$. Assume that the contrasts are all iid normal with 0 mean and variance τ^2 . Suppose that the absolute value of the contrasts are plotted vs. the quantiles of the half-normal distribution on half-normal probability paper (i.e., we construct a half-normal probability plot). Based on the previous discussion of Daniel's half-normal plots, we would expect that the contrasts would all lie on a straight line.

Now suppose that the effect sparsity principle holds and that, relative to n , only a small number of effects are active. In this case, we would expect the active contrasts to lie off of the straight line defined by the inert contrasts. Typically, one would use the probability plot to visually select the active contrasts.

The drawback of this method lies in the subjectivity of determining which contrasts lie far enough from the straight line to be deemed active. In order to reduce the subjectivity of this procedure, it would be useful to employ a statistic that will mirror the appearance of the half-normal plot in a quantitative manner and provide some decision rule to assess the significance of the contrasts.

As part of a simulation study of censored data in fractionated factorial experiments, Hamada and Wu (1991) present a method that attempts to mimic the behavior of an experimenter subjectively viewing a half-normal plot in a quantitative manner. They did this so they could use a simulation study to evaluate the performance of their proposed method for handling censored data in this context. It would be impossible to perform such a simulation study without some automatic procedure

for picking active contrasts and the authors desired the procedure to mimic what is done in practice with the half-normal plot.

The method proposed by Hamada and Wu uses a forward selection procedure, based on the R^2 -statistic of successive regressions of the contrasts and the quantiles of the half-normal distribution. In this manner, the authors attempt to quantitatively measure the degree to which the contrasts lie away from the straight line.

Background

Recall the simple linear regression model

$$u_i = \beta_0 + \beta_1 v_i + \epsilon_i \quad (2.1)$$

where,

u_i is the value of the response of the i^{th} trial

β_0 and β_1 are regression coefficients

the v_i 's are fixed, known constants

ϵ_i is a random error term such that $\epsilon_i \sim N(0, \tau^2)$, where ϵ_i and ϵ_j are uncorrelated and $\text{cov}(\epsilon_i, \epsilon_j) = 0$

$i = 1, 2, \dots, n$.

Typically, β_0 and β_1 are not known and are estimated using the method of least squares. That is, the function

$$S = \sum_{i=1}^n (u_i - (\beta_0 + \beta_1 v_i))^2$$

is minimized with respect to β_0 and β_1 , and estimates of β_0 and β_1 are determined.

The degree of linear association between the dependent variable and the independent variable in the aforementioned model is usually measured with the R-square statistic,

$$R^2 = \frac{s_{uv}^2}{s_u^2 s_v^2} \quad (2.2)$$

where, s_{uv}^2 is the square of the sample covariance of u and v , and s_u^2 and s_v^2 are the corresponding sample variances of u and v .

When the model fits the data well, the R-square value should be close to 1. When there exists no linear association between the variables, u and v , the R-square value is close to 0.

The Method

In order to mimic the behavior of an individual using the half-normal probability plots, Hamada and Wu present a step-wise procedure to detect active effects in factorial experiments. The aim of this method is to provide an automatic method for interpreting the half-normal probability plots.

Suppose that the effect sparsity principle holds for an unreplicated factorial experiment. Then it is reasonable to expect the smallest $\frac{n}{2}$ (or $\frac{n+1}{2}$ if n is odd) contrasts to be inert.

Therefore, if we fit a regression line through the smallest $\frac{n}{2}$ (or $\frac{n+1}{2}$) contrasts and the quantiles of the half-normal distribution, we would expect the R-square value to be close to 1. In this context, we are fitting a regression line through the smallest points of a half-normal plot. The remaining points can be added one at a time and the R-square recalculated. If an inert contrast is added, then little change in the R-square value will be observed. However, if an active contrast is added, then a drop in the R-square will be observed because of the lesser degree of association between the variables.

Consequently, when a large drop in the R-square is observed from the last fitted line (the authors suggest a drop of 0.1), then this is taken as evidence of the presence of active contrasts. That is, the degree by which this contrast deviates from the

straight line defined by the other contrasts is large enough to identify this contrast as active. When such a drop occurs, the currently added contrast and the remaining larger absolute contrasts are identified as estimating active effects.

Example 2.2: Hamada and Wu's method is applied to the data in example 2.1. Let R_s^2 be the R^2 value before a contrast has been added and R_e^2 be the R^2 after a contrast has been added for a given step.

Contrast Added	R_s^2	R_e^2	Difference
5	0.872	0.925	-0.053
8	0.925	0.948	-0.023
3	0.948	0.860	0.088
9	0.860	0.709	0.151

Since the procedure stopped when $\hat{c}_9$ was added, then this contrast and all contrasts with larger absolute values are identified as active. Therefore, contrasts 1, 2, 4 and 9 are identified as estimating active effects.

2.2.2 Lenth's Method

In an attempt to identify active effects in unreplicated factorial experiments, Lenth (1989) proposed a method based on a simple estimator of the standard error of the inert contrasts and Student's t -statistic. Using an estimate of the standard error based on a trimmed median of the contrasts, Lenth proposes using an acceptance region to simultaneously test the significance of each contrast. This section begins by introducing the components of Lenth's method followed by the method itself.

Background

In this section, we will introduce and discuss Lenth's estimate for the standard error of the contrasts. The properties of this estimator, the pseudo standard error, are

discussed in detail by Dong (1993a, 1993b).

Two important results needed to introduce the motivation behind Lenth's estimator are now presented. The proofs of these results are lengthy and are not presented in this paper (Dong, 1993a).

Consider an iid sample of size n , $(x_1, x_2, \dots, x_n)$, where except for k_n elements, $\{x_i\}_{i=1}^n$ have distribution function $F(\cdot)$. Then, the sample $\{x_i\}_{i=1}^n$ is contaminated by k_n elements. Let ξ_{p_o} be the p_o^{th} quantile of F , and let ξ_{n,p_o} be the p_o^{th} quantile of the empirical distribution function $\hat{F}_n$. Also, assume that

$$k_n = o(n^{-1/2}) \text{ as } n \rightarrow \infty. \quad (2.3)$$

The first result of this section states that, under (2.3), ξ_{n,p_o} is a consistent estimator of ξ_{p_o} .

Theorem 2.2: (Dong, 1993a) Let $(x_1, x_2, \dots, x_n)$ be independent and suppose that, except for k_n elements, $\{x_i\}_{i=1}^n$ are iid with distribution function $F(\cdot)$. Where F is differentiable at $\xi_{p_o} = F^{-1}(p_o)$ with derivative $F'(\xi_{p_o}) = f(\xi_{p_o}) > 0$, let ξ_{n,p_o} be the p_o^{th} sample quantile of the empirical distribution function $\hat{F}_n$. Then when (2.3) holds,

$$\xi_{n,p_o} = \xi_{p_o} + [p_o - F_n(\xi_{p_o})]/f(\xi_{p_o}) + R_n$$

where $R_n = o_p(n^{-1/2})$ and

$$\sqrt{n}(\xi_{n,p_o} - \xi_{p_o}) \rightarrow N(0, [p_o(1 - p_o)]/f^2(\xi_{p_o}))$$

Proof: See Dong 1993a. □

Suppose that the sample $(x_1, x_2, \dots, x_n)$ is truncated. That is $a \leq x_i \leq b$ for $i = 1, 2, \dots, n$. Let $\xi_{p_o,a,b}$ be the p_o^{th} quantile of the truncated distribution, then

$$\xi_{p_o,a,b} = F^{-1}[p_o F(b) + (1 - p_o)F(a)],$$

and $\xi_{n,p_o,a,b}$ be the p_o^{th} quantile of the corresponding empirical distribution function.

Note that if $p_o = \frac{1}{2}$ then $\xi_{p_o,a,b}$ is the median of the truncated distribution and $\xi_{n,p_o,a,b}$ is the median of the corresponding empirical distribution function. The second result of this section states that $\xi_{n,1/2,a,b}$ is a consistent estimator of $\xi_{1/2,a,b}$.

Theorem 2.3: (Dong, 1993a) Suppose that the x_i 's are such that $\{x_i : a \leq x_i \leq b, -\infty \leq a < b \leq \infty\}$. Let $(x_1, x_2, \dots, x_n)$ be independent and suppose that, except for k_n elements, $\{x_i\}_{i=1}^n$ are iid with distribution function $F(\cdot)$, where F is differentiable at $\xi_{1/2,a,b} = F^{-1}\{1/2[F(a) + F(b)]\}$ and $f(\xi_{1/2,a,b}) > 0$. When (2.3) holds,

$$\sqrt{n}(\text{med}(x_i) - \xi_{1/2,a,b}) \rightarrow N(0, [F(b) - F(a)]/[4f^2(\xi_{1/2,a,b})]). \quad (2.4)$$

Proof: See Dong 1993a. □

These results can be applied in a factorial experiment framework so that an estimate of the standard error may be found. Consider an unreplicated factorial experiment with n effects $c_1, c_2, \dots, c_n$ and independent contrasts $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$. Assume that, relative to n , the number of active effects is small. That is, the effect sparsity principle holds. Let k_n be the number of active effects in the factorial experiment. Then the remaining $n - k_n$ are simply the inert effects. Suppose that the inert contrasts are *iid* $N(0, \tau^2)$, and that the effect sparsity principle can be represented as

$$k_n = o(n^{-1/2}) \text{ as } n \rightarrow \infty.$$

The author considers using two estimates for the standard error of the inert contrasts in order to identify active effects. The first estimate is used to screen out potentially active contrasts from the calculation of the second estimate of the standard error of the inert contrasts. It is then hoped that the second estimate will then be based on the inert contrasts and, perhaps, a few small active contrasts.

Define

$$s_o = \text{med}_{1 \leq i \leq n} a|\hat{c}_i|$$

where a is a constant to be determined later. For the inert contrasts, if $x_j = |\hat{c}_j|$, then x_j has a half-normal probability density function with density,

$$f(x_j) = \sqrt{\frac{2}{\pi\tau^2}} e^{-\frac{x_j^2}{2\tau^2}}.$$

Theorem 2.4: (Dong 1993a) Let $\{\hat{c}_i\}_{i=1}^n$ be the independent contrasts from an unreplicated factorial experiment. Suppose that, except for k_n elements, $\{\hat{c}_i\}_{i=1}^n$ are iid $N(0, \tau^2)$. Let μ be the median of the half normal distribution with pdf $f(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}}$, $x > 0$. Assume that (2.3) holds for k_n . Then,

$$\sqrt{n}(s_o - a\mu\tau) \xrightarrow{d} N(0, \frac{\pi e^{\mu^2} a^2 \tau^2}{8}).$$

Proof:

Let $x_i = \frac{|\hat{c}_i|}{\tau}$ for $i = 1, 2, \dots, n$. Then, except for k_n elements, $f(x_j) = \sqrt{\frac{2}{\pi}} e^{-\frac{x_j^2}{2}}$. Applying Theorem 2.2, with $p_o = 1/2$,

$$\sqrt{n}(\text{med}_{1 \leq i \leq n}(x_i) - \mu) \rightarrow N(0, \frac{\frac{1}{2}(1 - \frac{1}{2})}{(\sqrt{\frac{2}{\pi}} e^{-\frac{\mu^2}{2}})^2}).$$

Equivalently,

$$\sqrt{n}(\text{med}_{1 \leq i \leq n}(x_i) - \mu) \rightarrow N(0, \frac{\pi e^{\mu^2}}{8}).$$

See that,

$$\begin{aligned} a\tau \{ \sqrt{n}(\text{med}_{1 \leq i \leq n}(x_i) - \mu) \} &= \sqrt{n}(a \cdot \text{med}_{1 \leq i \leq n}(|\hat{c}_i|) - a\mu\tau) \\ &= \sqrt{n}(s_o - a\mu\tau). \end{aligned}$$

Therefore,

$$\sqrt{n}(s_o - a\mu\tau) \rightarrow N(0, \frac{\pi e^{\mu^2}}{8} a^2 \tau^2).$$

□

Therefore, s_o is a consistent estimator of $a\mu\tau$. This estimator will be used to screen out potentially active contrasts so that an estimator of τ may be determined.

Define the pseudo standard error (PSE) as

$$PSE = a \cdot \text{med}(|\hat{c}_i| : |\hat{c}_i| < 2.5s_o)$$

and,

$$\widehat{PSE} = a \cdot \text{med}(|\hat{c}_i| : |\hat{c}_i| < 2.5\tau).$$

Let

$$\xi = \xi_{-\infty, 2.5} = F^{-1}[(1/2)F(2.5)].$$

Theorem 2.5: (Dong, 1993a) Let $\{\hat{c}_i\}_{i=1}^n$ be the independent contrasts from an unreplicated factorial experiment. Suppose that, except for k_n elements, $\{c_i\}_{i=1}^n$ are iid $N(0, \tau^2)$. Assume that (2.3) holds for k_n . Then,

$$\sqrt{n}(\widehat{PSE} - a\xi\tau) \xrightarrow{d} N(0, \frac{\pi F(2.5)e^{\mu^2}}{8}a^2\tau^2).$$

Proof: Let $x_i = \frac{|\hat{c}_i|}{\tau}$ for $i = 1, 2, \dots, n$. Then, except for k_n elements, $f(x_j) = \sqrt{\frac{2}{\pi}}e^{-\frac{x_j^2}{2}}$. Applying Theorem 2.3 with $a = -\infty$ and $b = 2.5$,

$$\sqrt{n}[\text{med}(x_i) - \xi] \rightarrow N(0, F(2.5)\frac{\pi e^{\xi^2}}{8}).$$

Therefore,

$$a\tau\sqrt{n}[\text{med}(x_i) - \xi] \rightarrow N(0, a^2\tau^2F(2.5)\frac{\pi e^{\xi^2}}{8}),$$

and

$$\sqrt{n}[\widehat{PSE} - a\tau\xi] \rightarrow N(0, a^2\tau^2F(2.5)\frac{\pi e^{\xi^2}}{8}).$$

□

If we choose $a = 1/\xi$, then $\widehat{PSE}$ is a consistent estimator of τ . By simulation, $\xi \approx 0.66$, therefore we choose $a = 1.5$.

So, Lenth defines an initial estimate of the standard error, s_o ,

$$s_o = 1.5 \text{med}|x_i|. \quad (2.5)$$

The expected value of the median of the half normal distribution, $f(x) = \sqrt{2/\pi}\tau^2 e^{-\frac{x^2}{2\tau^2}}$, is about 0.675τ (Lenth, 1989). Therefore,

$$\mu \approx 0.675\tau.$$

Then, s_o is a consistent estimator for $1.5 \cdot 0.675\tau \approx 1.01\tau$.

Since the value of τ is unavailable $\widehat{PSE}$ may not be determined. Substituting s_o for τ in the definition of $\widehat{PSE}$, we get what Lenth terms the pseudo standard error, PSE . See that the s_o is expected to overestimate τ . Therefore, we hope that the PSE is based on the inert contrasts and, perhaps a few active ones.

The Method

Lenth proposes to use the t -statistic and the pseudo standard error to construct an acceptance region for the contrasts of unreplicated factorial experiments. That is, for unreplicated factorial experiments with n contrasts Lenth's method simultaneously tests

$H_0: c_i$ is inert,

vs.

$H_1: c_i$ is active

$\forall i = 1, 2, \dots, n.$

The author assumes that all contrasts are independent, the effect sparsity principle holds, and that each contrast is normally distributed with equal variance and

not necessarily the same mean. That is,

$$c_i \sim N(k_i, \sigma^2).$$

Lenth then suggests that an acceptance region for each contrast be constructed using Student's t -statistic. Recall, the acceptance region for testing whether the mean of a distribution, based on a sample of size n , is different from 0,

$$\bar{x} \in \left[0 - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1}, 0 + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1} \right],$$

where s is the sample standard deviation of $(x_1, x_2, \dots, x_n)$, and $t_{\frac{\alpha}{2}, n-1}$ is the $(\alpha/2)^{th}$ quantile of the t -statistic with $n - 1$ degrees of freedom. In similar fashion, Lenth proposes an acceptance region for the individual contrasts

$$[0 - PSE * t_{\frac{\alpha}{2}, df}, 0 + PSE * t_{\frac{\alpha}{2}, df}].$$

Define the *margin of error*,

$$ME = PSE * t_{\frac{\alpha}{2}, df}.$$

Then the acceptance region for each estimate at a level of $(1 - \alpha)$ is,

$$[0 \pm ME].$$

The degrees of freedom are estimated by assuming that the distribution of the PSE^2 is similar to that of s^2 , and fitting the empirical distributions of PSE^2 to scaled chi-squared distributions by matching the first two moments. The author determined that $df = n/3$ was a fairly good estimate of the degrees of freedom.

The above acceptance region considers the individual estimates only. The level of the test must be adjusted in order to test all contrasts simultaneously. Let A_i = the event that contrast c_i is covered by the above acceptance region $\forall (i = 1, 2, \dots, n)$, if all effects are inert. Since, by assumption, all $\hat{c}_i$ are independent, then $P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i)$. Therefore, in order to get an exact two-sided, $(1 - \alpha)$ acceptance region,

the significance of the simultaneous test should be

$$\gamma = 1 - \frac{1 + (1 - \alpha)^{\frac{1}{n}}}{2}.$$

To account for the fact that many comparisons are being performed at one time, the *simultaneous margin of error (SME)* is defined as,

$$SME = PSE \cdot t_{\gamma, df}, \quad (2.6)$$

and the simultaneous $(1 - \gamma)$ acceptance region is,

$$[0 \pm SME]. \quad (2.7)$$

The author recommends using the acceptance region of (2.6) to identify active contrasts. Therefore if a contrast lies outside this interval, we reject the null hypothesis and say that the contrast estimates an active effect. However, if a contrast lies in the interval $[0 \pm SME]$ then there is not enough evidence to reject the null hypothesis, and the contrast is said to be inert.

Example 2.3: Lenth's method is applied to the data in example 2.1 with $\alpha = 0.05$. The acceptance region based on the *SME* is

$$[0 \pm 5.87].$$

Therefore, based on the above acceptance region, contrasts 1 and 2 are identified as estimating active effects.

2.2.3 Dong's Method

The method proposed by Lenth to identify active contrasts uses the observed contrasts to calculate s_o . In turn, s_o is used to calculate the pseudo standard error. In essence, Lenth's method uses s_o to screen out contrasts that are active and of relatively large magnitude, and finds an estimate of the standard error based upon the

remaining inert contrasts (and possibly a few smaller active ones). Similar to Lenth, Dong (1993b) proposed using s_o to screen out active contrasts. However, a different estimate of the standard error of the inert contrasts is recommended. Based on this new estimate, an acceptance region may be constructed to test whether a contrast is active.

Background

Suppose $c_1, c_2, \dots, c_n$ are the effects of a factorial experiment. Let $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$ be the corresponding contrasts where each contrast is independently normally distributed with equal variance but not necessarily the same mean. That is,

$$c_i \sim N(\mu_i, \tau^2).$$

Without loss of generality, the inert contrasts are such that,

$$\hat{c}_l \sim N(0, \tau^2) \quad \forall l = 1, 2, \dots, m$$

where, $m = \#\{i : c_i \text{ is inert}\}$.

Then,

$$E(\hat{c}_l) = 0 \quad \forall l = 1, 2, \dots, m,$$

and

$$\begin{aligned} \text{Var}(\hat{c}_l) &= E\{(\hat{c}_l^2 - E(\hat{c}_l))^2\} \\ &= E(\hat{c}_l^2) = \tau^2 \quad \forall l = 1, 2, \dots, m. \end{aligned}$$

So, since the $\hat{c}_k$'s are iid, then

$$E\left(\sum_{l=1}^m \hat{c}_l^2\right) = m\tau^2.$$

Therefore, if the inert contrasts are known, then an unbiased estimate of the variance of the contrasts is,

$$s_1^2 = \frac{\sum_{l=1}^m \hat{c}_l^2}{m}.$$

The Method

In practice, of course, we do not know which of the effects are active or inert. In order to pool the inert contrasts to compute a good estimate of the standard error, the inert contrasts must first be identified. Dong suggests using s_o to screen out active contrasts in precisely the same way as Lenth. The remaining contrasts are used to calculate s_1 . That is,

$$s_o = 1.5 \text{med}_{1 \leq i \leq n} |\hat{c}_i|$$

and,

$$s_1^2 = \frac{\sum_{\{i: |\hat{c}_i| \leq 2.5s_o\}} \hat{c}_i^2}{m} \quad (2.8)$$

where $m = \#\{i : |\hat{c}_i| \leq 2.5s_o\}$.

Now that an estimate of the standard error exists, Dong, like Lenth, suggests testing each of the contrasts with a $(1 - \alpha)$ simultaneous acceptance region based on the t -statistic. As before, the combined significance is

$$\gamma = \frac{(1 + (1 - \alpha)^{\frac{1}{n}})}{2}.$$

Then, the acceptance region used to test whether a contrast is active or inert are,

$$[0 \pm s_1 t_{\gamma, m}].$$

If $\hat{c}_i \in [0 \pm s_1 t_{\gamma, m}]$, then c_i is said to be inert, otherwise it is said to be active.

Example 2.4: Dong's method is applied to the data in example 2.1 with $\alpha = 0.05$. The acceptance region based on the s_1^2 is

$$[0 \pm 3.49].$$

Therefore based on the above acceptance region, effects 1, 2, 4 and 9 are identified as active by Dong's method.

2.2.4 *Benski's Method*

An iterative method based on a test for normality and an outlier test was proposed by Benski (1989) to identify active contrasts in unreplicated factorial experiments. Benski's method combines Olsson's (1979) version of the Shapiro-Wilk test for normality with a test for the detection of outliers, under the assumptions of effect sparsity, underlying normality and independence of contrasts. This section will introduce the normality test, the fourth spread outlier test and Benski's method.

Background

The Shapiro-Wilk test for normality is now introduced. Consider an independent normal random sample. Suppose that it is desirable to test

H_0 : The sample is iid normal

vs.

H_1 : The sample is not iid normal.

If a half-normal probability plot is constructed, under H_0 one would expect the points to fall on a straight line. Furthermore, if the sample is in fact iid normal, then the correlation coefficient, or R-square value, between the ordered sample points and the quantiles of the normal distribution should be close to 1. Under H_1 , the plot should be non-linear and the R-square should be small.

Based on this notion, Shapiro and Wilk (1965) proposed a test statistic, W , for normality based on the covariances of the order statistics of the $N(0, 1)$ distribution. Modifications (Shapiro & Francia, 1972) to this statistic were proposed in order to extend its use to larger sample sizes. Olsson's (1979) version of the Shapiro-Wilk W test for normality presents modifications of the W test in a simple to perform procedure. The procedure is presented below.

Let $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ be an ordered sample. The modified statistic can be represented as,

$$W' = \frac{(\sum_{i=1}^n m_i x_{(i)})^2}{\sum_{i=1}^n m_i^2 \sum_{i=1}^n (x_i - \bar{x})^2},$$

where m_i is Blom's (1958) estimate of the i^{th} expected order statistic of the standard Normal distribution,

$$m_i = \Phi^{-1}\left(\frac{(i - \alpha)}{(n - 2\alpha + 1)}\right),$$

where

Φ^{-1} is the inverse of the standard Normal cumulative distribution function,

and

$$\alpha = \begin{cases} 0.2755 + 0.0724 \cdot \ln(n)^{0.4115} & \text{if } 1 < i < n \\ 0.2051 + 0.1315 \cdot \ln(n)^{0.2267} & \text{if } i = 1 \text{ or } i = n. \end{cases}$$

Also define,

$$\begin{aligned} A &= 1.032 = 0.184(0.1n)^{-0.545}, \\ B &= -0.508 + 2.077(0.1n)^{-0.491}, \\ C &= \frac{(W' - A)/B + 0.049}{0.028} - \ln(100). \end{aligned}$$

Then,

$$P_1 = \exp(C)$$

is the probability that a normal distribution would give a W' as low as the observed W' .

Benski's method proposes using W' in conjunction with a test for outliers to identify relatively large contrasts. The outlier test is now defined.

Testing the hypothesis that the mean of the observed effects is 0 can be somewhat misleading since, for example, two large opposite effects may not significantly effect the central positioning of the sample. Instead, the *fourth-spread test* for outliers to the underlying distribution (Hoaglin, 1983) may be a better indication of the presence of large effects, as it is robust with respect to a few extreme values.

Consider $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, an ordered sample of size n from some distribution. Define the *depth of the median* as

$$dm = \frac{n+1}{2},$$

and the *depth of fourths* as

$$dpf = \frac{\lfloor dm \rfloor + 1}{2},$$

where the brackets $\lfloor dm \rfloor$ represent the largest integer smaller than dm . Now, the *lower fourth* and *upper fourth* can be defined respectively as

$$F_L = x_{(dpf)},$$

and

$$F_U = x_{(dm+dpf-1)}.$$

Interpolation should be used if dpf or $dpf + dm$ is not an integer. The *fourth-spread* is

$$fspread = F_U - F_L.$$

For symmetric distributions, the fourth-spread is simply the inter-quartile range. Data points that lie outside the interval $[F_L - 1.5fspread, F_U + 1.5fspread]$ can be considered outliers.

Consider a very large sample from a normal distribution. Since the normal distribution is symmetric, the fourths F_L and F_U correspond to 25% and 75% quantile points respectively. Therefore the lower and upper fourths are $\mu - 0.675\sigma$ and $\mu + 0.675\sigma$ where σ is the standard deviation of the sample. The fourth-spread is then 1.350σ and the interval used to test for the presence of outliers is $[\mu - 2.7\sigma, \mu + 2.7\sigma]$.

For finite samples, it has been found (Hoaglin, Iglewicz & Tukey, 1981) that a conservative upper bound for the significance of this test as

$$P_2 = 0.00698 + \frac{0.4}{n}.$$

This is remarkably high, since for a sample of size n we can expect $0.4 + 0.00698n$ contrasts to be outside the interval (Hoaglin 1983).

The Method

Consider an unreplicated factorial experiment with n effects, $c_1, c_2, \dots, c_n$, and contrasts $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$. If all contrasts are iid normal with mean 0, then they are simply noise from a normal distribution. Suppose the effect sparsity principle holds, and, relative to n , there is a small proportion of active contrasts.

Benski's method proposes to perform two tasks. Firstly, the test of normality is applied to detect the presence of non-normal contrasts. If the null hypothesis is rejected, this is taken as evidence of the possibility of the presence of active contrasts. Secondly, the presence of outliers is tested using the fourth-spread test. A confidence interval based on a robust estimate of the spread is chosen over a confidence interval based on the sample standard deviation because the sample standard deviation is contaminated by the active contrasts.

Let $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$ be the n contrasts of a factorial experiment, and $\hat{c}_{(1)}, \hat{c}_{(2)}, \dots, \hat{c}_{(n)}$ be the corresponding ordered contrasts. Then Benski's method can be summarized as follows:

- 1) Perform the normality test and obtain the significance level P_1 .
- 2) If P_1 is small (say < 0.05), then perform the outlier test, and identify the outliers. Go to step 4. Otherwise step 3.
- 3) If P_1 is not small, then list all the contrasts that have been rejected. These are the significant contrasts. Stop.
- 4) The combined significance level can be estimated by using Fisher's combination procedure (Stephens, 1986),

$$P_c = 2 \ln \left(\frac{1}{P_1 P_2} \right).$$

- 5) Remove the largest absolute contrast from the sample and associate it with the significance P_c . Re-label the remaining contrasts and go to step 1.

It should be noted that the significance levels of each of the tests are conservative approximations. The significance level of the fourth-spread test is such that one should expect a relatively high number of false positives, thus indicating that this test alone is inappropriate for determining if a contrast is active or not.

The combined significance P_c is based on the fact that $2 \ln(\frac{1}{P_1 P_2})$ is approximately χ_4^2 . Again, this is conservative.

Example 2.5: Benski's method is applied to the data in example 2.1. The Shapiro-Wilk statistic, the significance of the Shapiro-Wilk test and the confidence interval based on the fourth spread are listed below for each step of the procedure.

Contrast Tested	W'	P_1	$\pm 2 \cdot f_{spread}$
2	0.611	< 0.001	3.25
1	0.840	0.019	3.00
4	0.848	0.029	2.50
9	0.810	0.014	2.63
3	0.950	0.570	2.00

The test stops because the significance of the Shapiro-Wilk statistic for contrast 3 is greater than 0.05. Therefore based on the results of applying Benski's method, effects 1, 2, 4 and 9 are identified as active.

2.2.5 *Le and Zamar's Method*

In this section, a forward step-wise procedure to detect active effects (Le and Zamar, 1993) is presented. The method consists of using a test statistic based on the ratio of two M-estimates of scale (a *scale-ratio* test statistic) and a forward search procedure.

Background

Given any distribution function F and parameter $\theta > 0$, we can define

$$F_\theta(x) = F(x/\theta).$$

Then $\{F_\theta : \theta > 0\}$ is the *scale parameter family* generated by F , and θ is the *scale parameter*.

If $X \sim F$, then $\theta X \sim F_\theta$. See that larger values of θ tend to give larger observations. Therefore, a scale parameter tends to give an indication of the magnitude, relative to 0, of the observation with the distribution F_θ (Staudte and Sheather 1990).

The test statistic proposed by Le and Zamar is based on the ratio of two estimates of scale. The estimates proposed belong to a particular class of estimators, M-estimators:

Definition 2.1: Let ρ be an even function and T be a robust estimate of location. The corresponding M-estimate of scale, S , is defined as the solution to

$$\frac{1}{n} \sum_{i=1}^n \rho \left(\frac{(\hat{c}_i - T)}{S} \right) = E(\rho(Z))$$

where,

$$Z \sim N(0, 1).$$

M-estimators tend to have some useful properties. Firstly, in the case of the estimators proposed by Le and Zamar, they are consistent estimators of the scale parameters and asymptotically normal. Secondly, often when the selected ρ -function is smooth and bounded, the M-estimators are robust with respect to the presence of outliers.

The Method

Consider an unreplicated factorial experiment with n effects, $c_1, c_2, \dots, c_n$, and contrasts $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n$. In order to detect the presence of an active effect, it is reasonable to test

$$H_0: c_1 = c_2 = \dots = c_n = 0$$

vs.

$$H_0: \exists c_i \neq 0.$$

The authors propose using two estimators of scale to help identify active contrasts. The two estimators of scale, S_1 and S_2 , have differing properties. S_1 will be a robust estimator of scale (i.e., robust in the sense that it is insensitive to the presence of outliers), while S_2 will be a non-robust estimator of scale. This is a useful combination of properties since, when no outliers are present, the ratio of these two estimators should be close to 1. However, when outliers are present, the non-robust estimator should differ considerably from the robust estimator of scale. It should be noted that in the presence of effect sparsity, an outlier is assumed to be an active contrast.

The test is now defined. Let S_1 and S_2 be two M-estimators of scale for the following ρ -functions respectively:

$$\rho_1 = \begin{cases} x^2 & \text{if } |x| < \iota \\ \iota & \text{otherwise} \end{cases} \quad (2.9)$$

$$\rho_2 = \rho_1 + \eta(x^4 - 6x^2). \quad (2.10)$$

The authors suggest that reasonable values for ι and η are $\iota = 0.9$ and $\eta = 0.11$. These values were determined by the authors via simulation.

The scale-ratio test statistic based on n contrasts is defined as:

$$R_n = S_2/S_1. \quad (2.11)$$

The null hypothesis, H_0 , is rejected for large values of R_n . A table of critical values determined by Monte Carlo simulation using 10,000 simulation runs from the standard Normal distribution is presented in Le and Zamar (1993) Table 1 gives the 0.05 significance level critical values.

In order to address the possibility of multiple outliers, a forward procedure is proposed. That is, based on the n contrasts, S_1 and S_2 are determined. If the null hypothesis is rejected, then the contrast with the largest absolute value is removed and the test is performed again for the remaining $n - 1$ contrasts. The procedure is repeated until the test is no longer significant, and the non-robust statistic is close in value to the robust statistic.

If we assume that the contrasts are normally distributed, then the inert contrasts will be normally distributed with mean 0. It can be shown that the limiting distribution of the normalized scale ratio statistic $n^{-1/2}(R_n - 1)$ is normal with mean 0 under the null hypothesis.

Consider the local alternative,

$$H_1: \hat{c}_1, \hat{c}_2, \dots, \hat{c}_n \sim (1 - \epsilon)N(0, 1) + \epsilon N(\mu_i, 1)$$

$$\text{where } 0 < \epsilon < 1, \mu_i = \theta_i/n^{1/8} \text{ for some constant } \theta_i, \text{ and } \sum \theta_i^4/n \rightarrow \theta^4.$$

Under H_1 , the limiting distribution of $n^{-1/2}(R_n - 1)$ is also normal. Moreover, under these conditions, the test is a Pitman efficient test.

Example 2.5: Le and Zamar's method is applied to the data in example 2.1. The tabulated 0.05 significance level reference value and the scale ratio test statistic are presented below.

Number of Contrasts n	Table Value	R_n
15	1.48	7.29
14	1.52	2.56
13	1.52	2.03
12	1.55	1.84
11	1.55	1.11

The test stops when the sample is no longer significant. Therefore based on the results of applying Le and Zamar's method, effects 1, 2, 4 and 9 are identified as active.

Chapter 3

Simulation Study

In this chapter, we will compare the performance of the various automatic methods for identifying active effects in unreplicated factorial experiments through a simulation study. The goal of the study is to assess the ability of each method to identify the correct number of active effects in the presence of zero, one, two or three active effects of varying sizes. We begin by examining the performance of each of the methods in the presence of zero, one, two, and three active effects with normally distributed yields.

Let H_0 be the case where no active effects are present. For comparison purposes, the same significance level, α , for all tests is desirable. Otherwise, it is not possible to compare the relative abilities of the methods to identify active effects. The significance levels of the tests under H_0 were arrived at by the various authors through very different means. We must, however, attempt to set the significance level of the methods at the same value in order to compare results.

For example, Benski's method uses a normality test and an outlier test to detect the presence of active effects. When an active effect is identified, it is removed from the sample and the test is re-applied to the remaining contrasts. However, the combined significance of the two tests changes with the number of contrasts being tested. The identification of the presence of active effects is done using the normality

test. Therefore, by setting the significance of the normality test at a desired level, we can set the significance level of the test in the presence of no active effects

Both Lenth's method and Dong's modification of Lenth's method suggest using intervals based on the t -statistic. The authors assume that the contrasts are independent and that the effect sparsity principle applies. The methods define α -level acceptance regions for each effect, c_i . That is,

$$[\pm \hat{\tau} \ t_{\alpha/2, df}],$$

where $\hat{\tau}$ is the estimate of the standard error for the method in question and $\alpha = 0.05$. If $\hat{c}_i \notin [\pm \hat{\tau} \ t_{\alpha/2, df}]$, then this is taken as evidence that the i^{th} effect is active. However, Lenth suggests that since all contrasts are being tested simultaneously, the combined significance level of the test is

$$\gamma = 1 - \frac{[1 + (1 - \alpha)^{1/n}]}{2},$$

where, n is the number of effects.

The method of Le and Zamar defines critical values for sample sizes 5 to 32, and for $\alpha = 0.01, 0.05, 0.10$ and 0.15 . The critical values were approximated by Monte Carlo simulation using 10,000 runs (Le and Zamar, 1993). Therefore, we can easily set the significance level of this method by using the appropriate table value.

The method of Hamada and Wu unfortunately does not attempt to imply a significance level since the test statistic is not compared to any theoretical distribution. Rather, active effects are identified when a *large* jump (arbitrarily set at 0.1) in the R^2 is observed. Therefore, we are forced to apply the method without any knowledge of the significance of the test.

Setting the significance level of a test is quite important in practice. When an experimenter is presenting results, the results are meaningless if no significance level is specified. Each of the authors give a procedure for applying the method to get an approximate significance level (except Hamada and Wu whose method gives a

recommended procedure). We anticipate practitioners will apply the methods discussed in this thesis in the same manner as recommended by the authors. Therefore, we will first compare the methods applied as recommended for a desired significance level of $\alpha = 0.05$.

The method of Le and Zamar was applied using tabulated critical values with a significance level of 0.05. Similarly, Benski's method uses a normality test to identify the presence of active effects. Again, an α -level of 0.05 was chosen so that when the observed significance is less than α (i.e., $P_1 < \alpha$), this is taken as evidence of the presence of active contrasts. In the spirit of testing each contrast at a level of $\alpha = 0.05$, the combined significance of both Lenth's method and the modification of Lenth's method was chosen to be $\alpha = [1 + (1 - 0.05)^{1/n}]/2$. These significance levels are used throughout this chapter. By performing the simulation of the methods in this manner, we can best study how the methods will be applied in practice.

All simulations are programmed in the S-Plus statistical computer language. The important functions used are the S-Plus random normal generator

rnorm(),

and the random exponential generator used in Section 3.3

rexp().

The S-Plus random seed used for the simulations is,

$$seed = (45, 38, 19, 41, 7, 3, 43, 20, 36, 27, 47, 3) \quad (3.1)$$

3.1 Simulation With Normal Yields (Uncalibrated)

3.1.1 No Active Effects

It is important for any method to perform well in the absence of active effects. The false detection of an active effect can lead to costly erroneous conclusions. To this

end, the performance of the five methods is observed when no active effects are present.

Each of the five methods were tested for a 2^4 (or, for the purposes of this paper, a 16 run) factorial experiment. The yields of the experiments were generated from the standard normal distribution. That is, for each experiment

$$y_i = \epsilon_i \quad \forall i = 1, 2, \dots, 16, \quad (3.2)$$

where $\epsilon_i \sim N(0, 1)$. The resulting estimated effects were calculated and the tests applied. This was repeated for 10,000 simulation runs.

The proportion of detections of zero, one, two, three and greater than three active effects was recorded (Table 2). In relation to the experimenter's desire for a significance level of $\alpha = 0.05$, Benski's method and the method of Le and Zamar perform particularly well. This is expected, but for different reasons. We anticipate Le and Zamar's method to perform quite well when no active effects are present since the α -level critical values are determined via Monte Carlo simulations when no active effects are present. On the other hand, Benski's method detects the presence of active effects using an α -level test for normality. Therefore, the observed results are a reflection of the performance of the Shapiro-Wilk test statistic's ability to detect normality.

The simulation results indicate that Hamada and Wu's method fails to achieve a 0.05 significance level. This is not a surprise since no significance level was specified. The method incorrectly identifies the model at a level of about 0.10.

Both Lenth's method and Dong's method propose to test each individual effect at a level of 0.05. The simulation results show that when no active effects are present, the actual significance levels of the tests are 0.025 and 0.089 respectively. Therefore, while Lenth's method performs conservatively, Dong's modification to Lenth's method falsely identifies the model at a rate of almost twice the anticipated amount. The differences in the observed significance of the two methods is not

entirely expected since they both use the same procedure to screen out potentially active contrasts. However, some of the erratic behavior could possibly be attributed to the use of the t -statistic. That is, if a replicated factorial experiment with no active effects is performed, the sample variance of a contrast, $\hat{\tau}_i^2$, can be found, and

$$\frac{(\hat{c}_i - 0)}{\sqrt{\hat{\tau}_i^2}} \sim t_{(n-1)}.$$

However, neither,

$$\frac{(\hat{c}_i - 0)}{\sqrt{PSE^2}},$$

which Lenth treats as a t -variate, nor,

$$\frac{(\hat{c}_i - 0)}{\sqrt{s_1^2}},$$

which Dong treats as a t -variate, actually have a t -distribution. This is because both $PSE/[\sigma^2/(n-1)]$ and $s_1^2/[\sigma^2/(n-1)]$ are clearly not distributed as a $\chi_{(n-1)}^2$. Therefore, using the t -distribution as a reference distribution may give dubious results and the predicted level of significance may have little meaning.

Hamada and Wu's method does not imply any significance level, but it can be said that the method, as applied, achieves an observed significance of about 0.1. Based on these results, we can anticipate that Lenth's method will be less sensitive to the presence of active contrasts, while both Dong's method and Hamada and Wu's method will be more sensitive when compared to the other two methods.

At this point, we could calibrate the methods so that they achieve a 0.05 significance level. However, rather than calibrate the significance levels of the Dong's, Hamada and Wu's and Lenth's methods, this simulation study will continue applying the methods in the manner suggested by the authors. The investigation will proceed in this manner because the procedures recommended by the authors are the same as will be applied by experimenters in practice. We consider calibration in a later section so that we may assess the performance of these methods when they are adjusted to have a 0.05 significance level.

3.1.2 One Active Effect

A simulation study was done to investigate the performance of the methods in the presence of one active effect. The aim of this study is to observe the methods' relative abilities to detect the active effect.

As was the case in the simulation in the previous section, the five methods are tested for a 16 run unreplicated factorial experiment. The yields of the experiment are generated from a standard normal distribution. However, a constant C is added to the appropriate yields to make the first factor active. That is,

$$y_i = C \cdot I_1 + \epsilon_i \quad (3.3)$$

where, $\epsilon_i \sim N(0, 1)$ and

$$I_1 = \begin{cases} 1 & \text{if the first factor is at the low level} \\ 0 & \text{if the first factor is at the high level.} \end{cases}$$

The estimated effects based on these yields are then computed and the tests applied. This is repeated for 10,000 simulation runs.

In order to observe the behavior of the tests, the simulation is performed for $C = 1, 2$ and 3 respectively. These values were chosen so that the methods' ability to detect a relatively small, moderate, and a large active effect can be observed. The proportion of times zero, one, two, three and greater than three active effects was detected were recorded for each method in Table 3.

When one active effect is present, we are primarily interested in observing the number of times each method exactly fits the correct model, under fits the model and over fits the model. Notice in Table 3(a) that when $C = 1$ (a small effect is present), Dong's method and the method of Hamada and Wu perform considerably better than the other methods in exactly identifying the correct model, while the other methods are quite conservative in their ability to identify the active effect. These results may be anticipated based on the simulation results of the previous

section since Hamada and Wu's method and Dong's method had higher significance levels when no active effects were present.

Though Hamada and Wu's method chooses the correct number of active effects at a greater rate than all other methods, except Dong's, it has a tendency to incorrectly identify a very large number of active effects. The false identification of active effects may not be a desirable property of a method since it may lead the experimenter to make costly recommendations based on the experiment.

When $C = 2$ (a moderate size active effect is present), the relative performance of the methods appears to be closer; Table 3(b). While as expected, both Hamada and Wu's and Dong's methods are still able to identify the correct number of active effects with the highest degree of success, considerable gains are seen in the performance of both Benski's method and Le and Zamar's method. As anticipated, Lenth's method is still quite conservative in its detection of the active effect and under fits the model at the greatest rate. Overall, Dong's method performs best, choosing the correct model 64% of the time. Similar to the case when $C = 1$, when $C = 2$, Hamada and Wu's method tends to over fit the model by identifying more than one active effect at the highest rate.

When $C = 3$ (a relatively large effect is present), all methods, with the exception of Lenth's method, display similar abilities in identifying the presence of the active effect; Table 3(c). In fact, unlike the first two cases, when $C = 3$, both Le and Zamar's method and Benski's method out perform Hamada and Wu's method. In this case, Dong's method still performs best in that it exactly identifies the correct model most often. Lenth's method, on the other hand, is still considerably more conservative than the other methods tested. In fact, Lenth's method is the only method in this case that has a tendency to underfit the model. Similar to the case when $C = 2$, when $C = 3$, Hamada and Wu's method falsely detects a much greater proportion of cases where there is more than one active effect.

3.1.3 Two Active Effects

In order to examine the behavior of the methods in the presence of two active effects, a simulation study was performed. Again, a 16 run factorial experiment was considered. The yields are generated from a standard normal distribution, and the active effects are created by adding constants, C and D , to the low levels of the first two factors of the experiment. That is,

$$y_i = C \cdot I_1 + D \cdot I_2 + \epsilon_i \quad (3.4)$$

where, $\epsilon_i \sim N(0, 1)$,

$$I_1 = \begin{cases} 1 & \text{if the first factor is at the low level} \\ 0 & \text{if the first factor is at the high level,} \end{cases}$$

and

$$I_2 = \begin{cases} 1 & \text{if the second factor is at the low level} \\ 0 & \text{if the second factor is at the high level.} \end{cases}$$

The contrasts based on these yields were then computed and the tests applied. This is repeated for 10,000 simulation runs.

So as to observe the behavior of the methods in various cases with two active effects (e.g., two small active effects, two large active effects, one small and one large active effects,..., etc.), the simulation was repeated for all possible combinations of C and $D = 1, 2$, and 3 . The proportion of detection of zero, one, two, three and greater than three active effects were observed. The results are summarized in Table 4.

When both active effects are small, all of the methods tend to underfit the model. The simulation results indicate that the methods will underfit or exactly fit the model if two moderate active effects are present. As one would expect, the rate of identification of the correct model increases as the size of the two contrasts

increases. Therefore, the rate of detection of two active contrasts was observed to be highest when $C = 3$ and $D = 3$.

The results suggest that Dong's method outperforms the other methods regardless of the size of the effects. Furthermore, Hamada and Wu's method performs well relative to the other methods when the effects are small. However the performance of this method, relative to the other methods, decreases as the effect sizes increase. Moreover, this method consistently finds more than two active effects at a greater rate than all other methods tested. Benski's method and Le and Zamar's method performed similarly in that they performed well for two moderate or large effects, but were unable to detect the presence of two smaller effects. Again, Lenth's method is the most conservative in its ability to identify the correct model. However, this method also over fits the model the fewest number of times.

3.1.4 *Three Active Effects*

The behavior of the five methods is also tested with three active effects present. A simulation study of a 16 run unreplicated factorial experiment is performed. The experimental yields are generated from a standard normal distribution, and active effects are created by adding the constants C , D and E to the low levels of the first three factors of the experiment. That is,

$$y_i = C \cdot I_1 + D \cdot I_2 + E \cdot I_3 + \epsilon_i \quad (3.5)$$

where $\epsilon_i \sim N(0, 1)$,

$$I_1 = \begin{cases} 1 & \text{if the first factor is at the low level} \\ 0 & \text{if the first factor is at the high level,} \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if the second factor is at the low level} \\ 0 & \text{if the second factor is at the high level,} \end{cases}$$

and

$$I_3 = \begin{cases} 1 & \text{if the third factor is at the low level} \\ 0 & \text{if the third factor is at the high level,} \end{cases}$$

The corresponding effect estimates, based on these yields, are then computed and the tests applied. This procedure is repeated for 10,000 simulation runs. The simulation was performed for all possible combinations of C , D and $E = 1, 2$ and 3 . The proportion of detection of zero, one, two, three and greater than three active effects were observed. The results are summarized in Table 5.

Inspection of the results show that when the size of effects is small, most of the methods are unable to detect any significant contrasts. In these cases, Tables 5(a) and 5(b), Hamada and Wu's method and Dong's method exactly identify the correct model a greater proportion of the time.

As the size of the effects grows, the rate of identification of the correct model increases. When the effects are relatively large, Tables 5(i) and 5(j), in terms of exactly identifying the correct model the methods can be ranked as Dong's method followed by Hamada and Wu's method, Benski's method, Le and Zamar's method and Lenth's method. Again it should be noted that while the Hamada and Wu method performs relatively well in these cases, the rate of detection of greater than three active effects is about twice that of the next closest method. Similar to the results of the previous simulations, Dong's method performs best at all levels. Lenth's method, as anticipated from the simulation results in the absence of active effects, is most conservative for each combination of C , D , and E .

3.1.5 Discussion

Overall, it is observed that when relatively small active effects are present, Dong's method and Hamada and Wu's method identify the correct model most often, although, the latter method falsely identifies active effects at a much greater rate. When large effects are present, all methods, except Lenth's method, perform relatively well. Again, Dong's method outperforms the other methods in identifying the correct number of active effects, regardless of the size and number of active effects. Lenth's method is considerably more conservative than the other four methods.

When comparing results between simulation studies, the impact of masking can be seen. Masking typically occurs when moderate, similar sized effects are present. Their contribution to the sample of contrasts increases the variability enough so that small effects and effects of moderate size are undetected. For example, compare the simulation results with two active effects, $C = 2$ and $D = 2$ in Table 4(d), with the results with three active effects, $C = 2$, $D = 2$ and $E = 2$ in Table 5(g). The addition of the third, moderate sized active effect to the latter simulation affected the variability of the sample of effects so that the rate of identification of no active effects increased dramatically even though there was in fact one more active effect.

Similarly, comparing the results with one active effect and $C = 3$ in Table 3(c) with the simulation results when three active effects are present, $C = 1$, $D = 1$ and $E = 3$ in Table 5(c), we see that the simulation results with three active effects show that the rate of detection of no active effects more than doubles in most situations when compared to the case of one active effect. It is then reasonable to conclude that masking has an adverse affect on all of the methods.

Some important results of the simulation study are summarized:

- 1.) When no active effects are present, Benski's and Le and Zamar's methods come closest to achieving the desired level of significance.

- 2.) When one small active effect is present Dong's method and the method of Hamada and Wu are able to identify the correct model most often. When the one active effect is large Dong's and Benski's methods are able to identify the correct model at the greatest rate.
- 3.) With two and three relatively small effects present, Dong's method and Hamada and Wu's method identify the correct model with the highest degree of success. When the effects are relatively large, Dong's method performs best in terms of identifying the correct number of active effects.
- 4.) Regardless of the number of active effects, Lenth's method is quite conservative.
- 5.) Hamada and Wu's method often outperforms many of the methods, particularly when the active effects are relatively small. However, this method consistently falsely identifies more active effects than are present at a much greater rate than the other methods.
- 6.) All methods are affected by masking.

3.2 Simulation With Normal Yields (Calibrated)

In the discussion of the simulation results when no active contrasts are present, it was noted that some of the methods fail to come close to the desired significance level of $\alpha = 0.05$. This section discusses how the methods can be calibrated so that a significance level of 0.05 may be achieved.

The simulation results of the five methods in the presence of no active effects show that only Benski's method and Le and Zamar's method are able to perform at a significance level of 0.05. The simulation study continued to use these methods despite the differing α 's because the manner in which the methods were applied in

the simulation is the same manner an experimenter will likely apply these methods in practice. Although, it is reasonable to ask whether the conclusions reached in the simulation study would be the same if Dong's method, Lenth's method and Hamada and Wu's method were each performing with a significance level of 0.05 under H_0 .

3.2.1 *Calibrating the Methods*

To examine the performance of Dong's, Lenth's and Hamada and Wu's at the same significance level, under the null hypothesis, as the other two methods, we must first calibrate these methods to achieve a 0.05 significance under H_0 . To this end, the simulation with no active effects is repeated with a new seed

$$seed = (53, 42, 9, 59, 29, 1, 39, 61, 21, 51, 59, 0). \quad (3.6)$$

This seed is used for calibration purposes only.

This time, instead of using the significance levels recommended by the authors, both Dong's and Lenth's methods are calibrated by repeating the simulation for various "significance levels" of the t -statistic until the tests achieve a 0.05 significance level respectively. By performing this procedure, the new significance level for each individual effect for Dong's method and Lenth's method are recommended to be 0.0208 and 0.1215 respectively. If these significance levels are used with $n = 15$ contrasts under our model with no active effects, then the methods achieve a 0.05 level of significance.

Similarly, Hamada and Wu's method identifies active effects based on an observed drop in the R^2 statistic. The *drop value* of 0.1 is arbitrary and does not imply a significance level. In order to calibrate this method so that a 0.05 significance level may be achieved, the simulation study with no active effects is re-run with the seed in (3.1), and the method of Hamada and Wu is applied. However, the simulation study is repeated for various drop values in the R^2 until a significance of 0.05 is achieved. By proceeding in this fashion, Hamada and Wu's method achieves a 0.05

significance level, when no active effects are present, if the method detects active effects when a drop of greater than or equal to 0.184 in the R^2 is observed. That is, under our model with no active effect, Hamada and Wu's method identifies active effects, with a significance level of 0.05, when a drop value of 0.184 or more is observed.

3.2.2 *Simulation Study for Calibrated Methods*

To investigate the performance of the calibrated methods, a small simulation study was performed. Rather than test the calibrated methods for all of the cases in the previous simulation studies, a subset of the more interesting cases is selected.

In the spirit of providing results comparable to the previous simulations, the calibrated methods were applied to a 16 run unreplicated factorial experiment. The yields were generated from a standard Normal distribution, and constants C , D , and E were added to the appropriate factor levels in order to include the desired number of active effects. Therefore, the yields were randomly generated from,

$$y_i = C \cdot I_1 + D \cdot I_2 + E \cdot I_3 + \epsilon_i \quad (3.7)$$

where $\epsilon_i \sim N(0, 1)$,

$$I_1 = \begin{cases} 1 & \text{if the first factor is at the low level} \\ 0 & \text{if the first factor is at the high level,} \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if the second factor is at the low level} \\ 0 & \text{if the second factor is at the high level,} \end{cases}$$

and

$$I_3 = \begin{cases} 1 & \text{if the third factor is at the low level} \\ 0 & \text{if the third factor is at the high level,} \end{cases}$$

The corresponding effect estimates were computed and the calibrated tests applied. This procedure was repeated for 10,000 simulation runs.

As mentioned in previous discussions, it is important for any method to perform well when no active effects are present. Therefore, the simulation was performed with yields generated from (3.7) with $C = 0$, $D = 0$ and $E = 0$. The methods were also tested in the presence of one large active effect. The simulation was performed with yields randomly generated from (3.7) with $C = 3$, $D = 0$ and $E = 0$.

In order to observe the degree of success these methods have when masking is anticipated, the simulation is run with a few moderate sized active contrasts. That is, the simulation is performed with yields generated from (3.7) with $C = 2$, $D = 2$ and $E = 0$. The simulation is also repeated, but in the presence of one more active effect of moderate size (i.e., $C = 2$, $D = 2$ and $E = 2$). Masking is examined further by applying the calibrated methods to a case with one large active effect and a few small ones. The yields are generated from (3.7) with $C = 1$, $D = 1$ and $E = 3$.

3.2.3 *Simulation Results*

The simulation studies for the cases described in the previous section were run. The results are presented in Table 6 and are summarized in this section. Included in Table 6 are the simulation results of Benski's and Le and Zamar's methods. This is done so that the reader may easily compare the simulation results of all methods performing at a 0.05 significance level.

When there are no active effects present (i.e., $C = 0$, $D = 0$ and $E = 0$), the calibrated methods achieve a significance level that is close to 0.05; Table 6(a). This is expected since the calibration of these methods was based on the same simulation with a different random seed. The simulation results when no active effects are present are summarized below:

- 1.) When no active effects are present, each of the three calibrated methods achieve a significance level very close to 0.05.
- 2.) As anticipated, the calibrated version of Lenth's method is more sensitive to large values than it was before calibration. In fact Lenth's method, when calibrated, seems to perform similarly to Le and Zamar's method. Hamada and Wu's method as well as Dong's method become more conservative when calibrated.

When one large active effect is present (i.e., $C = 3$, $D = 0$ and $E = 0$), the calibrated methods identify the correct model with greater success than when they were uncalibrated; Tables 6(a) and 3(c). Particular improvement is seen in Dong's method in that the correct model is identified more often and the proportion of detection of greater than one active effect is roughly halved. One drawback is that the method under fits the model at a greater rate. This is largely the result of making the test more conservative. The results for the case of one large active effect are now summarized:

- 1.) All of the methods detect the correct model more often when calibrated and Dong's method, in particular, detects the correct model most often.
- 2.) Dong's method falsely detects active effects at a smaller rate when calibrated, but also fails to identify the active effect at a greater rate.
- 3.) Lenth's method detects the correct model more frequently when calibrated, but still much less frequently than Dong's method.
- 4.) The calibration of Lenth's method reduces the failure to detect any effects dramatically, but the method falsely identifies active effects at twice the rate of the uncalibrated method and at a rate greater than Dong's method.

5.) The calibrated method of Hamada and Wu detects the correct model more often than its uncalibrated counterpart. However, this method, even after calibration, tends to overfit the model more frequently than the other methods.

6.) When compared to Benski's method and Le and Zamar's method, only Dong's calibrated method outperform's both in that it identifies the correct model most often and falsely detects active effects least often. Moreover, the calibrated Dong's method fails to detect an active effect less often than Benski's method, Le and Zamar's method and the other two calibrated methods.

The simulation results with two moderate size active effects (i.e., $C = 2$, $D = 2$ and $E = 0$) show that Dong's method identifies the correct model most often; Table 6(c). The results of the simulation with two moderate size active effects are summarized:

1.) Dong's method identifies the correct model more frequently than the other calibrated methods. However, it also selects the correct model less often than when it was uncalibrated and fails to detect any active contrasts at a greater rate; Table 4(d). Again this is a result of making the test more conservative.

2.) Lenth's method is able to detect the correct model more frequently when calibrated and is able to detect the presence of at least one of the contrasts more often. However, the rate of falsely identifying active contrasts has increased substantially.

3.) The method of Hamada and Wu was able to identify the correct model less often when calibrated. When calibrated, it under fit the model more frequently but consistently over fit the model at a greater rate than the other methods.

4.) When compared to Benski's method and Le and Zamar's method, only Dong's method is able identify the correct model more often while at the same time falsely detecting active effects at a lower rate. Moreover, Dong's method is able to identify at least one of the active effects with greater frequency than both Benski's method and Le and Zamar's method.

When there are three moderate sized active effects (i.e., $C = 2$, $D = 2$ and $E = 2$), we anticipate that masking will be present since it was observed when the methods were uncalibrated. This situation was selected so that we may examine the performance of these methods in the presence of masking. The results are now summarized:

- 1.) The calibration of Lenth's and Dong's methods make the simulation results appear almost identical; Table 6(d). The two methods identified the correct model about 20% of the time (Dong's method performed slightly better).
- 2.) Dong's method identifies the correct model less frequently than when uncalibrated; Table 5(g).
- 3.) Lenth's method is able to detect the active effects more often when calibrated. On the other hand, when calibrated, Lenth's method detects greater than three active effects at a rate two times that of Dong's method.
- 4.) The method of Hamada and Wu identified the correct model less often when calibrated. Moreover, it fails to detect any active effects over 68% of the time.
- 5.) With respect to Benski's method and Le and Zamar's method, when calibrated, both Dong's method and Lenth's method perform better in

that they identify the correct model most often, fail to detect at least one active effect least often and detect greater than three active effects with the smallest frequency.

6.) The impact of masking can readily be seen. The simulation results indicate that the failure to detect any of the active effects increases for each of the calibrated methods with the addition of a moderate sized active effect. For example, when there are three moderate size active effects present, Table 6(d), the calibrated method of Dong detects zero active effects at a greater rate than when there are only two moderate sized active effects present; Table 6(c).

The simulation with one large active effect and two small sized active effects (i.e., $C = 1$, $D = 1$ and $E = 3$) is chosen so that we may further observe the effects of masking on the calibrated methods. The results in Table 6(e) suggest that:

- 1.) Dong's method detects the correct model at a very low rate (less than 2% of the time). However, it detects at least one of the active effects most often.
- 2.) When calibrated, Lenth's method detects the correct model most often. In fact, when compared to the same situation in the previous simulation study, Lenth's method is the only method to exactly identify the correct model more often after calibration. However, this calibrated method fails to detect any of the effects greater than 10% more often than Dong's calibrated method.
- 3.) The method of Hamada and Wu falsely identifies zero active effects most often. At the same time, it tends to over fit the model more often than the other methods.
- 4.) All the methods, including both Benski's method and Le and Zamar's

method do not detect any of the active effects a large proportion of the time. This is due to the masking of active effects.

Discussion

When considering the calibrated methods for all cases in this simulation, Dong's method performs best. It is able to identify the correct model at least as often as its closest competitor in most cases and it falsely detects active effects at a rate less than or equal to the other methods in each simulation. Moreover, when more than one active effect is present, Dong's method is able to detect at least one of the active effects at a greater rate than the other calibrated methods. Thus it is less sensitive to masking. Comparing the calibrated and uncalibrated results, we conclude that Dong's method performs best and this was not merely due to having a different α .

The overall results of the simulation are summarized below.

- 1.) Dong's method, when calibrated, is able to outperform the other two calibrated methods in identifying the correct model.
- 2.) Calibrating Lenth's method improves its performance in terms of identifying the correct model, but relative to the other methods it is still quite conservative in many cases.
- 3.) When Hamada and Wu's method is calibrated, it falsely detects active effects less often than before calibration. However this method, whether calibrated or uncalibrated, tends to over fit the model at a very high rate.
- 4.) If we compare the results of the simulation study for the calibrated methods with the simulation without calibration, we see that Dong's method, both uncalibrated and calibrated, is able to exactly identify the correct model most frequently. Furthermore, it also performs best in that it tends to under fit and over fit the model at the lowest rate.

5.) Masking affects each method, whether calibrated or not. The ability of each method to identify the correct model, when three moderate active effects are present, is not very good. In this case, each of the methods detects zero active effect more often than the correct model is identified.

3.3 Simulation Study With Non-Normal Yields

The simulations thus far have considered only normally distributed yields. Though many of the methods assume normality, in practice we know nothing about the underlying distribution of the yields, and therefore, nothing is known about the underlying distribution of the contrasts. Consequently, it is also important that the methods perform well when the yields of the factorial experiment come from some other, non-normal distribution. To examine the performance of the methods when the underlying distribution is non-normal a simulation study was performed applying the five methods and the three calibrated methods. In the interests of brevity, the simulation was performed for the same subset of cases considered in the simulation study of calibrated methods.

3.3.1 *Simulation Study*

To examine the performance of the methods when the yields are non-normally distributed, a simulation study was performed. In the spirit of comparing these results with the previous results in this thesis, the methods were applied to a 16 run unreplicated factorial experiment. The yields of the experiments were generated from an exponential distribution with mean 1 and variance 1. Moreover, in the interests of comparability with previous results, the mean 1 was subtracted from the yields so that the simulations are based on ζ_i 's with mean 0 and variance 1. In addition, constants C , D , and E were added to the appropriate factor levels to include the desired number of active effects. That is, for the factorial experiment, the yields

were generated from,

$$y_i = C \cdot I_1 + D \cdot I_2 + E \cdot I_3 + \zeta_i \quad \forall i = 1, 2, \dots, 16 \quad (3.8)$$

where,

$$\zeta_i = X_i - 1,$$

X_i 's are iid $\exp(\theta)$ with $\theta = 1$,

$$I_1 = \begin{cases} 1 & \text{if the first factor is at the low level} \\ 0 & \text{if the first factor is at the high level,} \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if the second factor is at the low level} \\ 0 & \text{if the second factor is at the high level,} \end{cases}$$

and

$$I_3 = \begin{cases} 1 & \text{if the third factor is at the low level} \\ 0 & \text{if the third factor is at the high level,} \end{cases}$$

The corresponding effect estimates were computed and the five methods and three calibrated methods applied. This is repeated for 10,000 simulation runs. In order to compare these results with previous simulation results, the simulation is run for each of the cases discussed in Section 3.2.

3.3.2 *Simulation Results*

The proportion of identification of zero, one, two, three and greater than three active effects is observed based on 10,000 simulation runs for each of the cases. The results are presented in Table 7.

Inspection of the simulation results in Table 7 suggest the same conclusions as the simulations with normally distributed yields (Tables 2-6). For example, comparing

the simulation results with no active effects present, Table 7(a), to the results of the previous section, Table 6(a), we see little difference in the relative abilities of the methods. Overall, each of the methods tend to be more conservative when the yields are generated as in (3.8).

Discussion

The simulation study for non-normally distributed yields displayed some encouraging results. For instance, the results are similar to the simulation results with normally distributed yields. This is an especially important characteristic of a good method since it suggests the methods are robust to the normality assumption. Each of the methods, in some heuristic sense at least, assume that the yields are iid normally distributed when there are no active effects present. It then follows that the contrasts are also iid normally distributed. Using different properties of the normal distribution, the methods argue that no active effects will be identified in this case. On the other hand, when an active effect is present, it is hoped that these same properties will identify the presence of the active effect. The success of the methods in identifying the correct model with non-normal yields was, in part, explained by Daniel (1959) who noted that regardless of the underlying distribution of the yields, there is some sort of analog to the central limit theorem at work that make the effect estimates appear normally distributed. Hence the wide use of the half-normal probability plots, and the success of these methods.

The simulation results with non-normal yields are summarized below:

- 1.) When no active effects are present, all methods perform well, and Dong's method as well as the method of Le and Zamar come closest to achieving the 0.05 significance level. Of the calibrated methods, Dong's and Lenth's methods only miss the desired significance level by about 1.5% and falsely identify active effects at a lower rate than Dong's

method (uncalibrated).

2.) When one active effect is present, Dong's method and Dong's method when calibrated perform best in that they are able to identify the correct model most often, and falsely detect active effects at a fairly low rate with non-normal yields.

3.) When two moderate sized active effects are present, each of the methods detect the correct model more frequently than when the yields are normally distributed. Again, Dong's method, both calibrated and uncalibrated, seem to perform best.

4.) When three moderate sized active effects are present, all methods are affected by masking. However, in the non-normal case, the methods seemed to perform slightly better than the similar case with normally distributed yields. In this case, the two methods of Dong perform best in that they are able to identify at least one of the active effect most often.

5.) When there are three active effects, one large and the other two relatively small, none of the methods are able to identify the correct model with any success. In fact, the ability of each of the methods to detect the correct model in this case was not as good as when the yields were normally distributed. Again, Dong's methods performed best.

Chapter 4

A Permutation Test to Detect Active Effects

In this chapter we introduce the notion of the permutation distribution in factorial experiments. A new test to detect the presence of active effects will be presented (Loughin and Nobel, 1994).

Consider an unreplicated factorial experiment with yields $(y_1, y_2, \dots, y_{n+1})$ and contrasts $(\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)$. In order to determine the existence of active effects, it is reasonable to test, for each effect,

$H_0 : c_i \text{ is inert}$

vs

$H_1 : c_i \text{ is active.}$

Suppose that the null hypothesis, H_0 , is true for each $i = 1, 2, \dots, n$. Then the factor level settings have no effect on the observed yields of the experiment. That is, each observed value of the yield would have occurred, regardless of the factor level settings. Therefore, the observed contrasts are the result of one possible assignment of yields to runs of the experiment.

In this case, there are $(n + 1)!$ possible assignments of yields to runs of the experiment. The collection of all $(n + 1)!$ possible arrangements (or permutations) of the

yields is called the *permutation distribution*. It is the properties of this distribution that motivate the test for active effects in the following section.

4.1 A Permutation Test

This section will present a new test based on the permutation distribution of the ordered absolute contrasts (OAC) of a factorial experiment proposed by Loughin and Nobel (1994). A sketch of the authors' construction of the method is presented, followed by their suggested algorithm.

Suppose that $(c_1, c_2, \dots, c_n)$ are effects from an unreplicated factorial experiment with estimates $(\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)$. In order to identify active effects in factorial experiments, Birnbaum (1961) proposed an optimal scale-invariant Bayes decision rule for selecting contrasts with non-zero means from a factorial experiment. Without loss of generality, suppose that $|\hat{c}|_{(1)} < |\hat{c}|_{(2)} < \dots < |\hat{c}|_{(n)}$. When the number of active effects in a factorial experiment is one, Birnbaum's test statistic is,

$$\hat{\omega}_1 = \frac{|\hat{c}|_{(n)}}{\sum_{i=1}^n \hat{c}_i^2}. \quad (4.1)$$

Large values of $\hat{\omega}_1$ are taken as evidence of the presence of active effects.

Also, suppose, for the moment, that all contrasts are inert. Then, the association of the $(n+1)$ yields to the runs of the unreplicated factorial experiment was purely coincidence, and is just one of the possible $(n+1)!$ assignments of the yields to runs. Consider any other permutation of the yields. The effect estimates can be calculated for this permutation of the yields. Repeating in this manner, effect estimates $(\hat{c}_{1,j}, \hat{c}_{2,j}, \dots, \hat{c}_{n,j})$ can be calculated for each possible permutation $j = 1, 2, \dots, (n+1)!$. In turn, the ordered absolute contrasts, $(|\hat{c}|_{(1),j}, |\hat{c}|_{(2),j}, \dots, |\hat{c}|_{(n),j})$, can be determined for each permutation of the yields. The collection of all $|\hat{c}|_{(i),j}$, $j = 1, 2, \dots, (n+1)!$, forms the permutation distribution for the i^{th} OAC. Moreover, if $\hat{\omega}_1$ is calculated

for each of the $(n + 1)!$ permutations, then the permutation reference distribution of $\hat{\omega}_1$ can be found.

Letting G^* be the permutation distribution of $\hat{\omega}_1$, given the observed yields $(y_1, y_2, \dots, y_{n+1})$, then "a non-parametric version of the optimal invariant decision rule is to reject the null hypothesis at the α -level of significance and conclude that the largest OAC has non-zero mean if the observed significance level of the test statistic is no larger than α " (Loughin and Nobel, 1994). That is, the null hypothesis is rejected if $G^*(\hat{\omega}_1) \leq \alpha$. In practice the permutation distribution is estimated using a large number of random permutations of the original sample.

The authors note that Birnbaum's test statistic for testing the presence of two or more active effects is impractical. This is due to the complexity of the joint distribution of $(\hat{\omega}_1, \hat{\omega}_2, \dots, \hat{\omega}_n)$. Instead, it is recommended that the effect of the contrast with the largest absolute value be removed, and the test reapplied. That is, to remove the effect let $\tilde{y} = \hat{y} - \hat{c}_1 x_1$, where $\hat{y}$ is the vector of yields, and x_1 is the column of the design matrix, X , corresponding to the contrast with the largest absolute value.

The effect estimates can be re-calculated based on the new vector of yields, $\tilde{y}$, and the permutation test re-applied to this set of contrasts. See that, due to the orthogonality of the contrasts of the design matrix, $\tilde{c}_1 = 0, \tilde{c}_2 = \hat{c}_2, \dots, \tilde{c}_n = \hat{c}_n$. It should be noted, however, that by removing the contrast with the largest absolute value, the variability due to this contrast has also been removed in the calculation of $\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_n$. This is not the case with $\tilde{c}_{1,j}^*, \tilde{c}_{2,j}^*, \dots, \tilde{c}_{n,j}^*$, the contrasts of the j^{th} permutation of $\tilde{y}$. It can be shown that the variability of the i^{th} contrast derived from a permutation of the yields differs by a factor of $n/(n - 1)$ (see Appendix).

In order to correct for the change in variability, let

$$\tilde{c}_{i,j}^* = \left(\frac{n}{n-1} \right)^{1/2} \tilde{c}_{i,j}^*.$$

Applying this adjustment to the observed significance based on the permutation

distribution,

$$p = 1 - [G^*(\hat{\omega}_2)]^{\frac{n-1}{n}}.$$

In general,

1. Compute $(\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)$ from $(y_1, y_2, \dots, y_{n+1})$, and order the contrasts and columns of the design matrix, $x_1, x_2, \dots, x_n$, to correspond to $|\hat{c}|_{(1)} < |\hat{c}|_{(2)} < \dots < |\hat{c}|_{(n)}$.

2. At step s compute $\hat{\omega}_s$, and obtain $\tilde{y}_s = \hat{y} - \hat{c}_n x_n - \hat{c}_{n-1} x_{n-1} - \dots - \hat{c}_{n-s+1} x_{n-s+1}$, where $\tilde{y}_1 = \tilde{y}$.

3. Select a large number, B , and repeat B times:

(a) Obtain $\tilde{y}_s^*$ through a random permutation of $\tilde{y}_s$.

(b) Compute $\tilde{c}_{(s)}^*$ from $\tilde{y}_s^*$.

(c) Obtain $|\tilde{c}^*|_{(n)}$ from $\tilde{y}_s^*$.

(d) Compute

$$\hat{\omega}_s^* = \left(\frac{n}{n-s+1} \right)^{1/2} \frac{|\tilde{c}^*|_{(n)}}{\sum_{i=1}^n \tilde{c}_i^{*2}}.$$

4. Compute the observed significance level (OSL),

$$p_s = 1 - \left[\frac{\# \hat{\omega}_s^* < \hat{\omega}_s}{B} \right]^{\left(\frac{n-s+1}{n} \right)}.$$

5. Repeat steps 2-4 for as many contrasts as desired.

The authors discuss two methods of applying the method. Firstly, it could be applied in a step-down manner. That is, if at any iteration s , the contrast with the largest absolute value is not significant, then that contrast and all remaining contrasts are identified as estimating inert effects.

The authors also discuss, and recommend, a step-up procedure. The significance, p_i , is calculated for all $i = 1, 2, \dots, n$ contrasts. All contrasts larger than or equal to the smallest contrast with an observed significance less than α (the actual significance level of the test) are considered to be active. In this manner, the method attempts to address the masking of significance due to a number of active effects of the same magnitude. The authors claim that this setup avoids the masking problem, but they only apply the method to a single data set.

At this point the reader has noticed that no assumptions have been made regarding the underlying distribution of the yields or contrasts. The other methods discussed in this thesis have used, to varying degrees, an assumption of normality. Since, in practice, nothing is typically known about the underlying distribution of the yields or contrasts, it is desirable to employ a non-parametric method.

The permutation method presented by the authors estimates the permutation distribution of Birnbaum's test statistic by taking a large number of permutations of the yields. This procedure is quite computer intensive since the permutation distribution is approximated for every contrast tested.

4.2 Simulation Study of the Permutation Method

In order to assess this method's ability to identify active contrasts in unreplicated factorial experiments, a simulation study was performed. The aim of the simulation is to observe the proportion of times the correct model is selected and to compare these results to the simulations of the other five methods and the calibrated methods. The simulation was performed for both the step-down and step-up permutation procedures.

In the spirit of providing comparable results, the permutation methods are tested for an unreplicated 16 run factorial experiment. The yields were generated from a standard Normal distribution, and constants C , D , and E are added to the appro-

priate levels in order to include the desired number of active contrasts. Therefore, the yields were randomly generated from,

$$y_i = \epsilon_i + C \cdot I_1 + D \cdot I_2 + E \cdot I_3 \quad (4.2)$$

where, $\epsilon_i \sim N(0, 1)$,

$$I_1 = \begin{cases} 1 & \text{if the first factor is at the low level} \\ 0 & \text{if the first factor is at the high level,} \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if the second factor is at the low level} \\ 0 & \text{if the second factor is at the high level,} \end{cases}$$

and

$$I_3 = \begin{cases} 1 & \text{if the third factor is at the low level} \\ 0 & \text{if the third factor is at the high level,} \end{cases}$$

The corresponding contrast estimates were then computed and the test applied. The random seed used is the same as was used in previous simulations. Rather than constructing the permutation distribution directly from the $16!$ possible assignments of the yields, the permutation distribution of the test statistic $\hat{\omega}_s$ is estimated by drawing $B = 1,000$ random permutations of $\mathbf{y}$. Due to time constraints, this procedure was repeated for 1,000 simulation runs instead of the 10,000 simulation runs used in the previous simulations.

To investigate this method's ability to choose the correct model, the simulation was performed for the same cases as in Sections 3.2 and 3.3. In order to compare these simulation results with those of the previous sections, a significance level of 0.05 was used.

4.2.1 *Simulation Results*

The proportion of detection of zero, one, two, three and greater than three active contrasts were recorded for both the step-down and step-up procedure. The results are presented in Table 8. These should be compared to the results presented in Tables 2-6. Particular emphasis should be placed on Table 6 where the results of the methods calibrated at a 0.05 significance level under the null hypothesis are presented.

The simulation results, with no active effects present (i.e., $C = 0$, $D = 0$, and $E = 0$), Table 8(a), are summarized below:

- 1.) The step down procedure comes closest to achieving the desired significance level (0.056). In the cases where the null hypothesis was rejected, the step-down procedure had very little tendency to falsely detect greater than one active contrast.
- 2.) The step-up procedure also comes close (but not as close) to achieving the significance level of 0.05. Like the step-down procedure, when active contrasts are detected, this procedure will tend to falsely detect only one.
- 3.) The step-up procedure incorrectly detects more than one active contrast at a greater rate than the step-down procedure. This is expected since when a contrast is deemed inert the step-down procedure will stop, while the step-up procedure continues to entertain the possibility that this contrast and others may, in fact, be active.
- 4.) If we compare these results to the simulation study of the other five methods and the two calibrated methods, Tables 2(a) and 6(a), when no active contrasts are present, then both the step-down and step-up procedures perform reasonably well with respect to identifying the correct model at a rate close to $\alpha = 0.05$.

When one large active effect is present (i.e., $C = 3$, $D = 0$, and $E = 0$), Table 8(b), both the step-down and the step-up procedures identify the correct model with a relatively high degree of success. We see that:

- 1.) Both the step-up and step-down procedures identify the correct model over 90% of the time (92.7% and 93.7% respectively).
- 2.) As expected, the step-up procedure will falsely detect active effects at a slightly greater rate than the step-down procedure.
- 3.) When comparing these results to the simulation results of other methods, Tables 3(c) and 6(b), under the same conditions, we see that both the step-down procedure and the step-up procedure perform better than the other methods in identifying the correct model. Only Dong's method performs nearly as well in this case.
- 4.) Both permutation methods falsely detect active contrasts at a relatively low rate when compared to the simulation results of the other methods.

The simulation was performed with two moderate sized active effects present (i.e., $C = 2$, $D = 2$, and $E = 0$); Table 8(c). The results are now summarized:

- 1.) When two moderate sized active contrasts are present only the step-up procedure identifies the correct model a majority of the time (58%).
- 2.) Even though there are two active contrasts present, the step-down procedure detects no active contrasts at a rate of three times greater than the step-up procedure.
- 3.) If we compare these results to the simulation results of the previous sections, Tables 4(d) and 6(c), the step-down procedure displays similar behaviour to the other methods in detecting the correct number of active

contrasts and also wrongly detecting no active contrasts. The step-up procedure is able to perform better than all of the methods in this case. Each of the five methods and the calibrated methods detect no active contrasts at a rate of greater than three times that of the step-up procedure. This demonstrates the step-up permutation method's ability to partially avoid masking.

4.) The step-up permutation method tends to over fit the model at about the same rate as methods that perform with a 0.05 significance level under the null hypothesis; Table 6(c). The step-down procedure displays very little tendency to falsely identify active contrasts in this case.

In order to observe the ability of the permutation methods to overcome the masking effect, the simulation is run with three moderate sized effects (i.e., $C = 2$, $D = 2$, and $E = 2$); Table 8(d). The simulation results are presented below:

- 1.) When three moderate sized effects are present, the ability of the step-up procedure to overcome the masking effect is displayed. While the step-down procedure detects no active effects almost 60% of the time, the step-up procedure detects no active effects only 24% of the time.
- 2.) The step-up and step-down procedures exactly identify the correct model at a rate of .398 and .039 respectively. So, compared to the step-down procedure, the step-up procedure performs strikingly well.
- 3.) If we now compare these results to the simulation results of the five methods and the calibrated results of previous sections, Tables 5(g) and 6(d), we see that the step-up procedure performs better than the other methods in identifying the correct model. In addition, the step-up method performs considerably better at detecting at least one of the active contrasts.

4.) Comparing the simulation results with two moderate sized contrasts with the simulation results with three moderate sized contrasts, we see that all of the methods fail to detect the presence of any active contrasts at a greater rate when the additional moderate contrast is added. That is, all methods are affected to some degree by the masking of contrasts. However, the step-up procedure is able to detect the presence of at least one of the active contrasts at a much greater rate. Therefore, although the step-up procedure is affected, to some degree, by masking, it is able to perform better than the other methods by a considerable margin when the masking of contrasts occurs.

The sensitivity of the methods to masking is further tested in the presence of one large contrast and two small ones (i.e., $C = 1$, $D = 1$ and $E = 3$); Table 8(e). The results are summarized:

- 1.) While the step-up procedure detected the correct model about 10% of the time, the step-down procedure was only able to detect the correct model about 3.5% of the time.
- 2.) The step-down model identified no active contrasts 55% of the time compared to only about 5% for the step-up procedure.
- 3.) If we compare these results to those for the simulation results of the five methods and the calibrates results in Tables 5(c) and 6(e), we see that the step-down procedure performs quite poorly compared to the other methods. The step-up method, on the other hand, performs much better than the other methods. With the exception of the calibrated method of Hamada and Wu, the step-up permutation method detects the correct model at a rate of about twice that of its closest competitor. Furthermore, it detects two of the three active contrasts at a rate of about

7 times that of the other methods. In addition, the step-up procedure has little tendency to over fit the model.

Discussion

Based on the simulations considered, the permutation method appears to provide improvement over the five methods considered in this thesis. When there is no masking present, the permutation methods perform as well or better than the other methods. However, when masking is present, the step-up procedure overcomes this obstacle with much greater success than the other permutation method and the methods introduced in Sections 2 and 3.2.

The results of this simulation are now summarized:

- 1.) Both the step-up and step-down procedure perform well when no active contrasts are present. The step-down procedure comes closest to achieving the significance level of $\alpha = 0.05$.
- 2.) When one active contrast is present, both methods identify the correct model at a rate higher than the other five methods and the calibrated versions. The step-up identifies more than one active contrast more often than the step-down method, but at a rate similar to the other five methods.
- 3.) When two moderate sized active contrasts are present, the step-up procedure performs better at overcoming the masking of active contrasts. The step-up procedure out performs all of the five methods dramatically in this case.
- 4.) When three moderate size active contrasts are present, both permutation methods are affected by masking, but the step-up procedure is affected to a much smaller degree.

5.) When there are three active contrasts, one large and the other two relatively small, both permutation methods have difficulty identifying the correct model. In fact, the step-down procedure detects no active contrasts a majority of the time. The step-up procedure performs best in that it is able to detect two or three active contrasts almost 90% of the time.

6.) The simulation results of the five methods and the calibrated methods show that none of the methods perform nearly as well as the step-up permutation method in overcoming masking for the situations addressed in this thesis.

4.3 Revisiting Dong's Method

The simulation study of the permutation method reveals some encouraging results. The step-up permutation method appears to be able to perform almost as well, or better, than all methods in all cases considered in this thesis. This method has particular success in overcoming the masking problem. It is interesting that the step-down algorithm does not perform nearly as well in the presence of masking. One might hypothesize that the success of the step-up permutation method in overcoming masking may be entirely attributed to the step-up nature of the procedure, and not to the method itself. If this is the case, then perhaps the other methods considered in this paper would benefit from application in a step-up manner.

To investigate whether a step-up adaptation may help the other methods detect active effects in unreplicated factorial experiments, a simulation study was performed using Dong's method and a step-up procedure. The simulation study considers Dong's method since, based on the simulation results in Chapter 3, Dong's method is the most successful of the five automatic methods considered.

The combined procedure of Dong's method and the step-up algorithm is as

follows:

- 1.) Compute $(\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)$ from $(y_1, y_2, \dots, y_{n+1})$, and order the contrasts and columns of the design matrix, $x_1, x_2, \dots, x_n$, to correspond to $|\hat{c}|_{(1)} < |\hat{c}|_{(2)} < \dots < |\hat{c}|_{(n)}$.
- 2.) Apply Dong's method, and test for the presence of active contrasts.
- 3.) At step s compute $\tilde{y}_s = \hat{y} - \hat{c}_n x_n - \hat{c}_{n-1} x_{n-1} - \dots - \hat{c}_{n-s+1} x_{n-s+1}$, where $\tilde{y}_1 = \hat{y}$.
- 4.) Repeat steps 2-3 for as many contrasts as desired.
- 5.) Let q be the last step in which an active effect was identified. Then there are q active effects.

Experimental yields were generated as in (4.2). This was repeated for 10,000 simulation runs for each of the cases discussed in the previous section. In addition, the simulation was applied to Dong's method in both the calibrated and uncalibrated cases. The results are presented in Table 9.

The results of the simulation are summarized:

- 1.) When $C = 0$, $D = 0$, and $E = 0$, Table 9(a), both the calibrated method and the uncalibrated method, applied in a step-up manner, are much less conservative. In this case, the methods perform much worse since they incorrectly identify active effects at a much higher rate. In fact, the calibrated method and the uncalibrated method falsely detect greater than three active effects at a rate of 0.1427 and 0.1946 respectively.
- 2.) When one relatively large active effect is present, Table 9(b), both methods exactly identify the correct model less often when applied in the step-up adaptation; Tables 3(c) and 6(b). Moreover, these methods

applied in the step-up manner display an alarming tendency to falsely detect many active effects.

3.) The simulation results with two moderate sized active effects (i.e., $C = 2$, $D = 2$, and $E = 0$) again suggest that the step-up permutation method performs better than Dong's methods applied in a step-up manner; Tables 9(c) and 8(c). The step-up permutation method identifies the correct model considerably more often than the two methods considered in this simulation. Again, the two methods, applied in a step-up manner, tend to over estimate the model at a very high rate.

4.) Applying Dong's methods in a step-up manner to the case where three moderate sized active effects are present (i.e., $C = 2$, $D = 2$, and $E = 2$) does not help overcome the masking effect; Table 9(d). Dong's method, whether uncalibrated or calibrated applied in a step-up manner, perform poorly compared to the step-up permutation method, Table 8(d), in selecting the correct model. In both cases they falsely identify greater than three active effects more than 25% of the time.

5.) If two small active effects are present and one large one (i.e., $C = 1$, $D = 1$, and $E = 3$), we see that both of Dong's methods identify the correct model less often when compared to the step-up permutation method; Tables 8(e) and 9(e). In addition, they tend to falsely detect active effects at a much greater rate.

6.) By applying Dong's methods in a step-up manner, we see in all cases that the methods will tend to falsely detect active effects at a greater rate than the permutation method.

7.) Based on the results of this simulation, applying Dong's method in a step-up manner does not help it overcome masking, and its overall performance is not as good as the step-up permutation method.

Chapter 5

Conclusions

Frequently, practitioners use 2^k factorial experiments to investigate some aspect of a product or process. Based on the experimental results, experimenters can recommend changes to improve the product or process. Due to constraints on time or experimental material, the experiment may be unreplicated. In this case, the usual analysis of variance techniques may not be used, leaving the experimenter without any standard method of identifying the important factors.

To overcome this problem, Daniel (1959) proposed a graphical method to detect active effects in unreplicated factorial experiments. This method plots the contrasts versus the quantiles of the half-normal distribution. The practitioner identifies active contrasts based on this plot. The half-normal plots rely on the experimenter's judgement to select the active effects from the plot, thus exposing this method to subjective bias.

To address the problem of identifying active contrasts in unreplicated factorial experiments, a number of automatic methods have been proposed. Hamada and Wu (1991) present a method to identify active effects that attempts to mimic an experimenter's use of the half-normal plot. Their method uses the R^2 statistic of successive regressions of the contrasts and the quantile of the normal distribution. When a large jump in the R^2 is observed, this is taken as evidence of active effects.

Lenth (1939) and Dong (1993b) propose using a trimmed median of the effect estimates to screen out potentially active contrasts and calculate estimates of the standard error of the inert contrasts. They use their respective estimators, along with the t -statistic to construct an acceptance region for the effects. Using these intervals, active effects are identified.

Benski (1989) introduced a method that combines the Shapiro-Wilk test for Normality and a robust estimate of the spread to identify active effects. This method first tests for normality. If the normality test fails, then the fourth-spread test is applied. If an active contrast is detected, it is removed from the sample and the test is reapplied, otherwise, the test concludes.

A scale-ratio test was proposed by Le and Zamar (1993) to detect active effects. The test statistic is a ratio of two M-estimators of scale: one which is robust in the presence of outliers and one which is not. In this context, active contrasts are represented as outliers. When no active effects are present, the estimates should be similar in value and the ratio of the estimators should be close to 1. If there exists an active effect in the experiment, then the robust estimate would not be affected by its presence. However, the non-robust estimate of scale would be impacted by the presence of the active effect and the ratio of the two estimators should not be close to one. Large values of the ratio of the non-robust and the robust estimates of scale is taken as evidence of active effects.

To investigate the performance of these automatic methods, a simulation study was performed for many cases with normally distributed yields. When no active effects were present, all methods are able to identify the correct model a great majority of the time, but Dong's, Lenth's and Hamada and Wu's methods failed to achieve the desired 0.05 significance level. In a later section, these methods were calibrated to a 0.05 significance level and the methods reapplied to some of the more interesting cases of the first section.

When one small active effect is present, Dong's and Hamada and Wu's methods

are able to identify the correct model most often. As the size of the effect gets larger, the effectiveness of the other methods improved. When a large active effect is present, the simulation study suggests that Dong's method is able to identify the correct model most often.

The simulation study with two and three active effects presented some interesting results. Based on the simulation results, Dong's method appears to be most successful in choosing the correct model, while at the same time detecting at least one of the active effects and not falsely identifying effects as active. However, the most striking result of the simulations with multiple active effects and normally distributed yields is the impact of masking. When a few moderate sized effects are present, the success of all the methods in detecting the active effects drops considerably.

The simulation with no active effects indicated that Lenth's, Dong's and Hamada and Wu's methods were not performing with a 0.05 significance level. To observe the performance of these methods when they actually achieve the desired significance level, the methods were calibrated. Simulations were then performed for some of the more interesting cases where zero, one, two and three active effects were present. If these results are compared to the first simulation results, we see that the more sensitive methods (Dong's and Hamada and Wu's) become more conservative and the more conservative method (Lenth's) becomes more sensitive. The results also suggest that the two most successful methods are Dong's method, calibrated and uncalibrated, and that masking is still a factor.

To investigate the performance of all the methods, and calibrated methods, when the yields of the experiment are non-normal a simulation is performed. Interestingly, the methods perform slightly more conservatively, but with similar success as the case where the yields were normally distributed. This is a promising result since nothing is typically known about the underlying distribution of the yields. Again, Dong's method performs best, but like the other methods is adversely affected by the masking of effects.

To identify active contrasts in factorial experiments and overcome the masking of effects, Loughin and Nobel (1994) introduce a method using the permutation distribution of the yields. The authors suggest applying a permutation test to the largest effect and finding its significance. The effect of this factor is then removed and the test reapplied. They introduce two different ways to apply their test. Firstly, a step-down procedure, where the test concludes when a non-significant effect is found, is introduced. The authors recommend a step-up procedure that tests each of the contrasts. The smallest significant contrast and all those greater are identified as estimating active effects. The aim of the step-up procedure is to overcome the masking of effects. The drawback of the permutation methods is that they are quite computer intensive.

To compare the performance of the permutation methods with the other methods already studied, a simulation study was performed. The simulation results of the step-up permutation method were very encouraging. The performance of this method was similar to the best of the methods when masking was not present. However, when masking is present, the step-up permutation outperforms the other methods by a considerable margin.

The success of the step-up permutation method in the presence of masking is striking. Intuitively, one may wonder if the step-up procedure may help the performance of other methods when a few moderate sized effects are present. To this end, the more successful methods, Dong's methods, were reapplied in a step-up manner. The simulation results of Dong's methods applied in a step-up procedure suggest that this application does not allow the methods to overcome masking. Regardless of whether there is masking present, the simulation results indicate that these methods applied in this manner will falsely detect active contrasts at a very high rate.

In closing, the simulation results of this thesis appear to indicate that the step-up permutation method is a very promising method for detecting active contrasts in

unreplicated factorial experiments. While performing as well or better as the other automatic methods, this method is able to overcome the masking of active effects with considerably more success, and in addition makes no assumptions on the form of the underlying distribution of the yields.

Table 1: 0.05 Significance Level Critical Values for Le and Zamar's Method

Sample Size	Critical Value
5	1.78
6	1.76
7	1.73
8	1.71
9	1.59
10	1.56
11	1.56
12	1.55
13	1.52
14	1.52
15	1.48
16	1.45
17	1.45
18	1.45
19	1.45
20	1.43
21	1.42
22	1.42
23	1.42
24	1.41
25	1.40
26	1.39
27	1.38
28	1.37
29	1.37
30	1.37
31	1.37
32	1.36

Table 2: Simulation Results With No Active Contrasts

Method	0	1	2	3	> 3
Benski	0.9577	0.0327	0.0059	0.0016	0.0021
Dong	0.9114	0.0650	0.0164	0.0045	0.0027
Hamada and Wu	0.8971	0.0496	0.0162	0.0085	0.0286
Lenth	0.9747	0.0159	0.0051	0.0019	0.0024
Le and Zamar	0.9428	0.0398	0.0084	0.0043	0.0047

2 of / de 2

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

Table 3: Simulation Results With One Active Contrast

Table 3(a) - $C = 1$					
Method	0	1	2	3	> 3
Benski	0.9087	0.0749	0.0111	0.0041	0.0012
Dong	0.8251	0.1400	0.0234	0.0079	0.0036
Hamada and Wu	0.8255	0.1082	0.0199	0.0120	0.0344
Lenth	0.9533	0.0328	0.0076	0.0031	0.0032
Le and Zamar	0.8875	0.0855	0.0148	0.0055	0.0067
Table 3(b) - $C = 2$					
Benski	0.4768	0.4849	0.0295	0.0062	0.0026
Dong	0.2989	0.6348	0.0503	0.0120	0.0040
Hamada and Wu	0.3291	0.5676	0.0480	0.0166	0.0387
Lenth	0.7546	0.2242	0.0132	0.0041	0.0039
Le and Zamar	0.4497	0.5070	0.0281	0.0079	0.0073
Table 3(c) - $C = 3$					
Benski	0.0625	0.8931	0.0344	0.0070	0.0030
Dong	0.0150	0.9142	0.0546	0.0120	0.0042
Hamada and Wu	0.0269	0.8661	0.0510	0.0172	0.0388
Lenth	0.3520	0.6257	0.0142	0.0042	0.0039
Le and Zamar	0.0692	0.8799	0.0329	0.0092	0.0088

Table 4: Simulation Results With Two Active Contrasts

Table 4(a) - $C = 1, D = 1$					
Method	0	1	2	3	> 3
Benski	0.9075	0.0653	0.0192	0.0049	0.0031
Dong	0.8018	0.1422	0.0407	0.0100	0.0053
Hamada and Wu	0.8094	0.0992	0.0372	0.0141	0.0401
Lenth	0.9445	0.0359	0.0114	0.0081	0.0044
Le and Zamar	0.8744	0.0878	0.0228	0.0081	0.0069
Table 4(b) - $C = 1, D = 2$					
Benski	0.5822	0.3413	0.0624	0.0091	0.0050
Dong	0.3607	0.5048	0.1069	0.0166	0.0074
Hamada and Wu	0.4292	0.4035	0.0978	0.0233	0.0472
Lenth	0.7853	0.1754	0.0278	0.0061	0.0054
Le and Zamar	0.5234	0.3929	0.0595	0.0140	0.0102
Table 4(c) - $C = 1, D = 3$					
Benski	0.1187	0.7893	0.0766	0.0095	0.0057
Dong	0.0311	0.8213	0.1230	0.0172	0.0074
Hamada and Wu	0.0704	0.7453	0.1137	0.0230	0.0476
Lenth	0.4166	0.5390	0.0324	0.0065	0.0055
Le and Zamar	0.1044	0.7951	0.0748	0.0149	0.0108

Table 4(d) - $C = 2, D = 2$

Method	0	1	2	3	> 3
Benski	0.4430	0.2090	0.3149	0.0258	0.0073
Dong	0.2570	0.2445	0.4524	0.0365	0.0096
Hamada and Wu	0.3739	0.1007	0.4202	0.0489	0.0563
Lenth	0.5697	0.1745	0.1387	0.0111	0.0060
Le and Zamar	0.4169	0.2365	0.3052	0.0273	0.0150

Table 4(e) - $C = 2, D = 3$

Benski	0.1295	0.3684	0.4654	0.0282	0.0085
Dong	0.0393	0.3071	0.6053	0.0387	0.0096
Hamada and Wu	0.1309	0.2012	0.5592	0.0520	0.0567
Lenth	0.3719	0.3879	0.2221	0.0119	0.0062
Le and Zamar	0.1153	0.3909	0.4473	0.0305	0.0160

Table 4(f) - $C = 3, D = 3$

Benski	0.0413	0.0770	0.8386	0.0331	0.0100
Dong	0.0120	0.0291	0.9073	0.0419	0.0097
Hamada and Wu	0.0406	0.0090	0.8380	0.0554	0.0570
Lenth	0.2558	0.2029	0.5222	0.0127	0.0064
Le and Zamar	0.0535	0.0954	0.7988	0.0357	0.0166

Table 5: Simulation Results With Three Active Contrasts

Table 5(a) - $C = 1, D = 1, E = 1$					
Method	0	1	2	3	> 3
Benski	0.9263	0.0465	0.0160	0.0079	0.0033
Dong	0.8128	0.1279	0.0386	0.0142	0.0065
Hamada and Wu	0.8269	0.0808	0.0337	0.0176	0.0410
Lenth	0.9461	0.0325	0.0120	0.0050	0.0044
Le and Zamar	0.8826	0.0787	0.0209	0.0090	0.0088
Table 5(b) - $C = 1, D = 1, E = 2$					
Benski	0.7029	0.2312	0.0434	0.0161	0.0064
Dong	0.4649	0.4003	0.0963	0.0295	0.0090
Hamada and Wu	0.5483	0.2806	0.0818	0.0383	0.0510
Lenth	0.8229	0.1348	0.0270	0.0091	0.0062
Le and Zamar	0.6085	0.3000	0.0590	0.0180	0.0145
Table 5(c) - $C = 1, D = 1, E = 3$					
Benski	0.2159	0.6989	0.0587	0.0187	0.0081
Dong	0.0666	0.7734	0.1184	0.0324	0.0092
Hamada and Wu	0.1426	0.6657	0.0990	0.0410	0.0517
Lenth	0.4973	0.4527	0.0338	0.0100	0.0062
Le and Zamar	0.1600	0.7267	0.0756	0.0222	0.0155
Table 5(d) - $C = 1, D = 2, E = 2$					
Benski	0.5939	0.1624	0.1833	0.0483	0.0121
Dong	0.3552	0.2365	0.3198	0.0731	0.0154
Hamada and Wu	0.4994	0.0807	0.2595	0.0929	0.0675
Lenth	0.7244	0.1468	0.0984	0.0220	0.0084
Le and Zamar	0.5079	0.2098	0.2099	0.0496	0.0228

Table 5(e) - $C = 1, D = 2, E = 3$					
Method	0	1	2	3	> 3
Benski	0.2399	0.3787	0.3080	0.0590	0.0144
Dong	0.0858	0.3426	0.4690	0.0867	0.0159
Hamada and Wu	0.2423	0.1954	0.3891	0.1046	0.0686
Lenth	0.4527	0.3431	0.1699	0.0254	0.0089
Le and Zamar	0.1805	0.3984	0.3366	0.0603	0.0242
Table 5(f) - $C = 1, D = 3, E = 3$					
Benski	0.0991	0.1262	0.6847	0.0737	0.0163
Dong	0.0279	0.0598	0.7952	0.1009	0.0162
Hamada and Wu	0.1086	0.0107	0.6914	0.1197	0.0696
Lenth	0.3244	0.2060	0.4301	0.0305	0.0090
Le and Zamar	0.0910	0.1174	0.6881	0.0774	0.0261
Table 5(g) - $C = 2, D = 2, E = 2$					
Benski	0.5519	0.0927	0.1129	0.2187	0.0238
Dong	0.3302	0.1600	0.1714	0.3101	0.0283
Hamada and Wu	0.4816	0.0189	0.0615	0.3375	0.1005
Lenth	0.6492	0.1410	0.1032	0.0925	0.0141
Le and Zamar	0.4863	0.1435	0.1146	0.2167	0.0389
Table 5(h) - $C = 2, D = 2, E = 3$					
Benski	0.2824	0.2181	0.1711	0.2987	0.0297
Dong	0.1156	0.2362	0.2149	0.4033	0.0300
Hamada and Wu	0.3225	0.0646	0.0862	0.4233	0.1034
Lenth	0.4097	0.2880	0.1573	0.1297	0.0153
Le and Zamar	0.2067	0.2754	0.1830	0.2930	0.0419

Table 5(i) - $C = 2, D = 3, E = 3$

Method	0	1	2	3	> 3
Benski	0.1372	0.0941	0.2956	0.4383	0.0348
Dong	0.0456	0.0685	0.2994	0.5542	0.0321
Hamada and Wu	0.1874	0.0050	0.1491	0.5517	0.1068
Lenth	0.2905	0.1858	0.3018	0.2056	0.0163
Le and Zamar	0.1163	0.1122	0.2952	0.4306	0.0457

Table 5(j) - $C = 3, D = 3, E = 3$

Method	0	1	2	3	> 3
Benski	0.0678	0.0283	0.0816	0.7817	0.0406
Dong	0.0261	0.0227	0.0439	0.8723	0.0350
Hamada and Wu	0.0644	0.0003	0.0068	0.8176	0.1109
Lenth	0.2126	0.1361	0.1699	0.4638	0.0176
Le and Zamar	0.0184	0.0528	0.0731	0.7433	0.0494

Table 6: Simulation Results For Calibrated Methods

Table 6(a) - $C = 0, D = 0, E = 0$					
Method	0	1	2	3	> 3
Benski	0.9577	0.0327	0.0059	0.0016	0.0021
Dong	0.9502	0.0381	0.0076	0.0029	0.0012
Hamada and Wu	0.9561	0.0219	0.0067	0.0031	0.0122
Lenth	0.9457	0.0331	0.0119	0.0046	0.0047
Le and Zamar	0.9428	0.0398	0.0084	0.0043	0.0047
Table 6(b) - $C = 0, D = 0, E = 3$					
Benski	0.0625	0.8931	0.0344	0.0070	0.0030
Dong	0.0401	0.9213	0.0303	0.0050	0.0033
Hamada and Wu	0.0747	0.8771	0.0243	0.0067	0.0172
Lenth	0.1514	0.7890	0.0345	0.0090	0.0071
Le and Zamar	0.0692	0.8799	0.0329	0.0092	0.0088
Table 6(c) - $C = 2, D = 2, E = 0$					
Benski	0.4430	0.2090	0.3149	0.0258	0.0073
Dong	0.3632	0.2686	0.3438	0.0197	0.0047
Hamada and Wu	0.5829	0.0604	0.3105	0.0223	0.0239
Lenth	0.4592	0.2430	0.2581	0.0279	0.0118
Le and Zamar	0.4169	0.2365	0.3052	0.0273	0.0150

Table 6(d) - $C = 2, D = 2, E = 2$

Method	0	1	2	3	> 3
Benski	0.5519	0.0927	0.1129	0.2187	0.0238
Dong	0.4348	0.1679	0.1607	0.2210	0.0156
Hamada and Wu	0.6810	0.0072	0.0298	0.2341	0.0479
Lenth	0.4463	0.1679	0.1638	0.1929	0.0291
Le and Zamar	0.4863	0.1435	0.1146	0.2167	0.0389

Table 6(e) - $C = 1, D = 1, E = 3$

Benski	0.2159	0.6989	0.0587	0.0187	0.0081
Dong	0.1365	0.7633	0.0772	0.0179	0.0051
Hamada and Wu	0.2964	0.6033	0.0556	0.0212	0.0235
Lenth	0.2584	0.6322	0.1638	0.0760	0.0125
Le and Zamar	0.1600	0.7267	0.0756	0.0222	0.0155

Table 7: Simulation Results With Non-Normal Yields

Table 7(a) - $C = 0, D = 0, E = 0$					
Method	0	1	2	3	> 3
Benski	0.9692	0.0204	0.0038	0.0047	0.0019
Dong	0.9430	0.0395	0.0080	0.0068	0.0027
Hamada and Wu	0.9228	0.0343	0.0072	0.0111	0.0246
Lenth	0.9848	0.0088	0.0028	0.0022	0.0014
Le and Zamar	0.9573	0.0277	0.0060	0.0058	0.0032
Dong (calibrated)	0.9678	0.0208	0.0050	0.0047	0.0017
Hamada and Wu (calibrated)	0.9639	0.0170	0.0026	0.0056	0.0109
Lenth (calibrated)	0.9655	0.0202	0.0070	0.0046	0.0027
Table 7(b) - $C = 0, D = 0, E = 3$					
Benski	0.1103	0.8590	0.0205	0.0048	0.0054
Dong	0.0544	0.9029	0.0310	0.0065	0.0052
Hamada and Wu	0.0502	0.8725	0.0360	0.0088	0.0325
Lenth	0.3544	0.6326	0.0081	0.0022	0.0027
Le and Zamar	0.1282	0.8345	0.0255	0.0076	0.0032
Dong (calibrated)	0.0836	0.8924	0.0161	0.0046	0.0033
Hamada and Wu (calibrated)	0.0847	0.8785	0.0175	0.0043	0.0150
Lenth (calibrated)	0.1912	0.7753	0.0211	0.0068	0.0056

Table 7(c) - $C = 2, D = 2, E = 0$

Method	0	1	2	3	> 3
Benski	0.3785	0.1579	0.4371	0.0166	0.0099
Dong	0.2500	0.1775	0.5440	0.0206	0.0079
Hamada and Wu	0.2929	0.0909	0.5408	0.0341	0.0413
Lenth	0.6185	0.1511	0.2195	0.0071	0.0038
Le and Zamar	0.3945	0.1769	0.3992	0.0175	0.0119
Dong (calibrated)	0.3337	0.1819	0.4685	0.0112	0.0047
Hamada and Wu (calibrated)	0.4449	0.0535	0.4449	0.0163	0.0204
Lenth (calibrated)	0.4408	0.1842	0.3489	0.0173	0.0088

Table 7(d) - $C = 2, D = 2, E = 2$

Benski	0.4522	0.0688	0.0922	0.3647	0.0221
Dong	0.3082	0.1145	0.1176	0.4406	0.0191
Hamada and Wu	0.3706	0.0277	0.0523	0.4755	0.0739
Lenth	0.5968	0.1165	0.1001	0.1776	0.0090
Le and Zamar	0.4442	0.1081	0.0919	0.3286	0.0272
Dong (calibrated)	0.3905	0.1116	0.1155	0.3716	0.0108
Hamada and Wu (calibrated)	0.5263	0.0145	0.0239	0.3974	0.0379
Lenth (calibrated)	0.4258	0.1247	0.1296	0.3003	0.0196

Table 7(e) - $C = 1$, $D = 1$, $E = 3$

Method	0	1	2	3	> 3
Benski	0.2164	0.6597	0.0702	0.0443	0.0094
Dong	0.1017	0.6953	0.1278	0.0654	0.0098
Hamada and Wu	0.1349	0.6238	0.1135	0.0811	0.0467
Lenth	0.4659	0.4654	0.0439	0.0192	0.0056
Le and Zamar	0.1979	0.6622	0.0811	0.0445	0.0143
Dong (calibrated)	0.1610	0.6993	0.0932	0.0412	0.0053
Hamada and Wu (calibrated)	0.2707	0.5935	0.0615	0.0500	0.0243
Lenth (calibrated)	0.2776	0.5885	0.0809	0.0418	0.0112

Table 8: Simulation Results of Permutation Methods

Table 8(a) - $C = 0, D = 0, E = 0$					
Method	0	1	2	3	> 3
Step-up	0.931	0.055	0.012	0.000	0.002
Step-down	0.944	0.055	0.001	0.000	0.000
Table 8(b) - $C = 0, D = 0, E = 3$					
Step-up	0.023	0.927	0.038	0.009	0.003
Step-down	0.027	0.937	0.036	0.000	0.000
Table 8(c) - $C = 2, D = 2, E = 0$					
Step-up	0.079	0.309	0.580	0.025	0.007
Step-down	0.249	0.316	0.420	0.014	0.001
Table 8(d) - $C = 2, D = 2, E = 2$					
Step-up	0.241	0.200	0.130	0.398	0.031
Step-down	0.611	0.297	0.052	0.039	0.001
Table 8(e) - $C = 1, D = 1, E = 3$					
Method	0	1	2	3	> 3
Step-up	0.065	0.056	0.762	0.103	0.014
Step-down	0.555	0.065	0.344	0.035	0.001

Table 9: Simulation Results of Step-up Procedure for Dong's Method

Table 9(a) - $C = 0, D = 0, E = 0$					
Method	0	1	2	3	> 3
Calibrated	0.8176	0.0252	0.0088	0.0057	0.1427
Uncalibrated	0.7382	0.0396	0.0163	0.0113	0.1946
Table 9(b) - $C = 0, D = 0, E = 3$					
Calibrated	0.0325	0.7557	0.0346	0.0133	0.1639
Uncalibrated	0.0121	0.6938	0.0512	0.0223	0.2206
Table 9(c) - $C = 2, D = 2, E = 0$					
Calibrated	0.2883	0.1554	0.3180	0.0426	0.1957
Uncalibrated	0.1771	0.1313	0.3695	0.0606	0.2615
Table 9(d) - $C = 2, D = 2, E = 2$					
Calibrated	0.2846	0.0610	0.1067	0.2744	0.2733
Uncalibrated	0.1761	0.0489	0.0960	0.3213	0.3577
Table 9(e) - $C = 1, D = 1, E = 3$					
Calibrated	0.1123	0.5788	0.0734	0.0365	0.1990
Uncalibrated	0.0492	0.5286	0.1042	0.0530	0.2650

Figure 1. Half-Normal probability plot of data in example 2.1

Appendix

In this appendix, we sketch the authors' reasoning for the correcting for unequal variances between the observed and permuted contrasts as discussed in Loughin and Nobel (1994). The case where $s = 2$ is presented.

Let $\mathbf{y}$ be the vector of yields from an unreplicated 2^k factorial experiment. Also, let $\mathbf{y}^*$ be a random permutation of $\mathbf{y}$. Under the null hypothesis, the the association of the yields and the runs of the experiment is purely coincidental. It is therefore easily shown that,

$$\bar{y}^* = \bar{y}$$

and

$$E^*(\hat{c}_i^* | \mathbf{y}) = 0 \quad \forall i = 1, 2, \dots, n$$

where $E^*(\cdot | \mathbf{y})$ is the expectation over all permutations of $\mathbf{y}$. By the same notion, we see that the variability also does not change, that is

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i^* - \bar{y}^*)^2.$$

It can be shown that,

$$\sum_{i=1}^n (y_i - \bar{y})^2 = 2^{n-1} \sum_{i=1}^n \hat{c}_i^2.$$

Similarly,

$$\sum_{i=1}^n (y_i^* - \bar{y}^*)^2 = 2^{n-1} \sum_{i=1}^n \hat{c}_i^{*2}.$$

Hence,

$$\sum_{i=1}^n \hat{c}_i^2 = \sum_{i=1}^n \tilde{c}_i^{*2}.$$

Consider the case where $s = 2$. Suppose, without loss of generality, that $|\hat{c}_1| < |\hat{c}_2| < \dots < |\hat{c}_n|$. Let $E^*(\cdot|\tilde{\mathbf{y}})$ be the expectation over all permutations of $\tilde{\mathbf{y}}$, where

$$\tilde{\mathbf{y}} = \mathbf{y} - \hat{c}_n \mathbf{x}_n.$$

By the orthogonality of the design matrix, $\hat{c}_1 = 0$. Then, since $\hat{c}_n = 0$,

$$\begin{aligned} \sum_{i=1}^n (\tilde{y}_i - \bar{\tilde{y}})^2 &= 2^{n-1} \sum_{i=1}^n \tilde{c}_i^2 \\ &= 2^{n-1} \sum_{i=1}^{n-1} \tilde{c}_i^2. \end{aligned}$$

In addition, since permuting the yields does not change their variability, then

$$\sum_{i=1}^n (\tilde{y}_i - \bar{\tilde{y}})^2 = \sum_{i=1}^n (\tilde{y}_i^* - \bar{\tilde{y}}^*)^2.$$

Therefore,

$$2^{n-1} \sum_{i=1}^{n-1} \tilde{c}_i^2 = 2^{n-1} \sum_{i=1}^n \tilde{c}_i^{*2}.$$

Note that the upper summation limit for the permuted yields is n . This is because $\tilde{c}_n^* \neq 0$. This implies that the inherent variability of $\tilde{c}_i$, $\forall i = 1, 2, \dots, n-1$ is $n/(n-1)$ greater than that of $\tilde{c}_i^*$, $\forall i = 1, 2, \dots, n$. It is for this reason that the authors apply the correction for the variance.

References

- Benski, H.C. (1989). "Use of Normality to Identify Significant Effects in Factorial Designs". *Journal of Quality Technology*, 21, 174-178.
- Blom, G. (1958). *Statistical Estimates and Transformed Beta Variables*. John Wiley & Sons Inc, Toronto.
- Birnbaum, A. (1961). "A Multi-Decision Procedure Related to the Analysis of Single Degrees of Freedom". *Annals of the Institute of Statistical Mathematics*, 12, 227-236.
- Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). *Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building*. John Wiley & Sons Inc, Toronto.
- Box, G.E.P. and Meyer, R.D. (1986). "An Analysis for Unreplicated Fractional Factorials". *Technometrics*, 28, 11-18.
- Casella, G. and Berger, R.L. (1990). *Statistical Inference*. Wadsworth Inc., Belmont, Ca..

- Daniel, C. (1959). "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments". *Technometrics*, 1, 311-341.
- Dong, F. (1993a). "Asymptotic Properties of Quantiles for Truncated Contaminated Data". *Communications in Statistics - Theory and Methods*, 22(11), 3255-3261.
- Dong, F. (1993b). "On the Identification of Active Contrasts in Unreplicated Fractional Factorials". *Statistics Sinica*, 3, 209-217.
- Hamada, M and Wu, C.F.J. (1991). "Analysis of Censored Data From Highly Fractionated Experiments". *Technometrics*, 33, 25-37.
- Hoaglin, D.C. (1983). "Letter Values: A Set of Selected Order Statistics". In *Understanding Exploratory Data Analysis*. (D.C. Hoaglin, F. Mosteller, and J.W. Tukey, Eds.), John Wiley & Sons Inc, New York, NY.
- Hoaglin, D.C., Iglewicz, B. and Tukey, J.W. (1981). "Small Sample Performance of a Resistant Rule for Outlier Detection". *1980 Proceedings of the Computing Section*. American Statistical Association, Washington D.C..
- Lenth, R.V. (1989). "Quick and Easy Analysis of Experiments". *Technometrics*, 31, 469-473.
- Le, N.D. and Zamar, R.H. (1992). "A Global Test For Effects in 2^k Factorial Designs Without Replicates". *Journal of Statistical Computation and Simulation*, 41-54.

- Loughin, T.M. and Nobel, W. (1994). "A permutation Test for Effects in an Unreplicated Factorial Design". Technical Report, Dept. of Statistics and Statistical Laboratory, Kansas State University.
- Montgomery, D.C. (1991). *Design and Analysis of Experiments*. John Wiley & Sons Inc, New York.
- Olsson, D.M. (1979). "Analysis of Two-Level Factorial Experiments". *Journal of Quality Technology*, 14, 18-27.
- Quinian, J. (1985). "Product Improvement by Application of Taguchi Mehtods". *American Supplier Institute News*, 11-16.
- Shapiro , S.S. and Francia, R.S. (1972). "Approximate Analysis of Variance Test for Normality". *Journal of the American Statistical Association*, 67, 215-216.
- Shapiro , S.S. and Wilk, M.B. (1965). "An Analysis of Variance Test for Normality (Complete Samples)". *Biometrika*, 52, 591-611.
- Staudte, R.G. and Sheather, S.J. (1990). *Robust Estimation and Testing*. John Wiley & Sons Inc, New York, NY.
- Stephens, M.A. (1986). "Tests for the Uniform Distribution". In *Goodness of Fit Techniques*. (R.B. D'Agostino and M.A. Stephens Eds.), Vol. 68, Marcel Dekker Statistics Textbooks and Monographs, New York, NY.

END

0 1 - 0 6 - 9 5

FIN