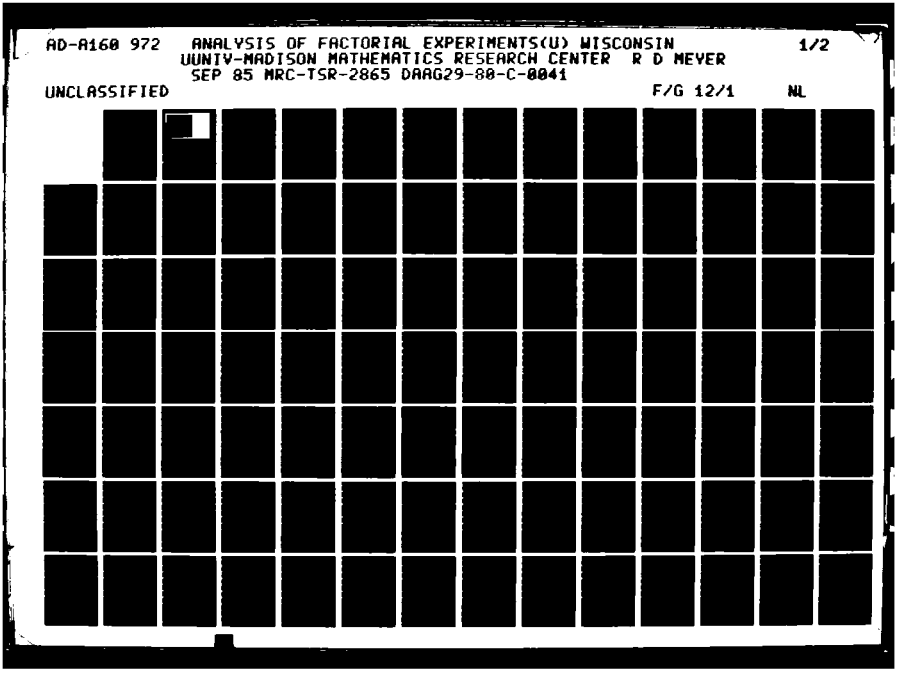
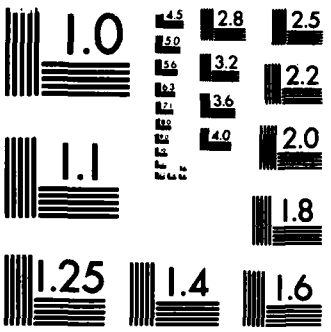

AD-A160 972 ANALYSIS OF FACTORIAL EXPERIMENTS(U) WISCONSIN 1/2
UUNIV-MADISON MATHEMATICS RESEARCH CENTER R D MEYER
SEP 85 MRC-TSR-2865 DAAG29-80-C-0041
UNCLASSIFIED F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

MRC Technical Summary Report #2865

ANALYSIS OF FACTORIAL EXPERIMENTS

R. Daniel Meyer

AD-A160 972

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

September 1985

(Received August 28, 1985)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, D. C. 20550

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

ANALYSIS OF FACTORIAL EXPERIMENTS

R. Daniel Meyer

Technical Summary Report #2865
September 1985

ABSTRACT

Fractional factorial designs have long been a key tool for the industrial statistician. They have received renewed attention recently due to the movement toward quality improvement sparked by the success of the Japanese in penetrating markets formerly dominated by western countries.

Fractional factorial designs are usually not replicated, so that it is not possible to estimate error variance in the usual way from repeat observations. Past methods of analysis have rested on an implicit hypothesis of effect sparsity, that most of the estimated effects measure only noise. Formalization of this hypothesis leads to a Bayesian analysis in which the posterior probability that an effect is active can be computed. A similar approach can be employed to obtain the posterior probability that a particular experimental factor is active. These probabilities are readily interpreted by graphical means, and provide a straightforward method for identifying active contrasts and active factors. In addition, the model is extended to the situation where there are possible outliers in the original observations. The posterior probability that an effect is active can be computed taking into account the possibility of bad values, and the posterior probability that an observation is bad can be computed taking into account that the identity of active effects is unknown.

AMS (MOS) Subject Classifications: 62A15, 62K15, 62N10

Key Words: Fractional factorial, unreplicated, active effects, Bayesian inference, outliers

Work Unit Number 4 (Statistics and Probability)

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041, by the National Science Foundation Grant No. DMS-8420968, and by the Vilas Trust of the University of Wisconsin-Madison, and aided by access to the research computer of the University of Wisconsin Statistics Department.

SIGNIFICANCE AND EXPLANATION

When many variables must be studied in an experiment, it is often too expensive to include repeat experimental runs for purposes of measuring the magnitude of noise. In the past, statistical analysis of such unreplicated experiments has relied on an assumption of effect sparsity, that the measured effect of most variables could be attributed to noise. Formalizing this assumption in probabilistic terms leads to a method of analysis in which the probability that the effect of a particular variable is too large to attribute to noise is computed. The probabilities associated with each of the variables are readily interpreted when presented graphically. The analysis can be extended to the case when it is thought that there may be misspecified values among the experimental observations.

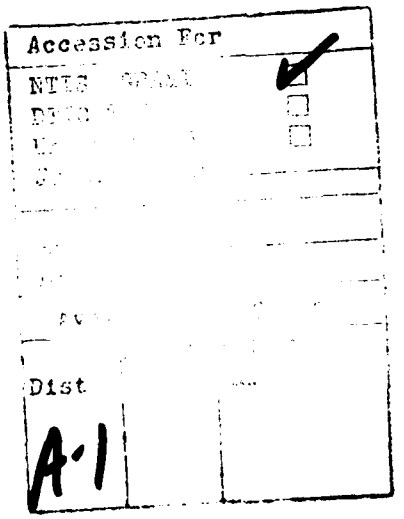


TABLE OF CONTENTS

R. Daniel Meyer

CHAPTER 1

INTRODUCTION

1.1. Motivation

The use of statistical methods in industrial improvement of quality and productivity has always been an important topic. It has received renewed attention recently due in part to the application of these methods by the Japanese and their success in penetrating markets formerly dominated by the United States (see, e.g., Deming, 1982).

A problem frequently encountered in this area is to identify from among many variables, those which are responsible for large changes in the quality characteristics of a particular process. Statistically designed experiments, in particular fractional factorial designs, are a key tool in providing an economical solution to this problem.

1.2. Fractional Factorial Designs

The possible value of fractional factorial designs in industry seems to have been first recognized by Tippett (1934) (see also Fisher, 1966, p. 88). To discover the cause of difficulty in a cotton-spinning machine, he successfully screened five factors, each having five levels, in just 25 runs: a 125th fraction of a 5^5 factorial. A general framework for fractional factorials was described by Finney (1945). More general orthogonal array designs were introduced by Plackett and Burman (1946) and Rao (1947).

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041, by the National Science Foundation Grant No. DMS-8420968, and by the Vilas Trust of the University of Wisconsin-Madison, and aided by access to the research computer of the University of Wisconsin Statistics Department.

At the preliminary stages of an investigation, a two-level fractional factorial is very useful as a screening design. While Plackett and Burman (1946) gave a fairly complete enumeration of two-level designs involving a moderate number of runs, the 2^{k-p} fractional factorials are an especially useful subset and a thorough description of them was given by Box and Hunter (1961). Because the Hadamard product of any two columns of a 2^{k-p} design gives another column of the design, the confounding structure is much simpler than for the general two-level orthogonal array. (The Hadamard product of two columns is defined to be a column with i th element equal to the scalar product of the i th elements of the original columns).

For ease of illustration, I will limit discussion here to two-level designs. It is assumed that the design matrix $\mathbf{X}$ is a $n \times n$ orthogonal matrix of $\pm 1's$ such that $\mathbf{X'X}=\mathbf{XX'}=n\mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix. The first column $\mathbf{x}_0$ of $\mathbf{X}$ is a column of 1's, and some or all of the remaining columns $\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}$ are assigned to experimental variables; -1 denoting the low or nominal level and +1 denoting the high or alternate level. At the completion of the experiment the $n \times 1$ vector $\mathbf{y}=(y_1, \ldots, y_n)'$ becomes available.

Typically, a linear model is employed for describing the observations from a two-level factorial experiment. At the screening stage of an investigation, it is often hoped that a first order model in main effects only will be adequate. This is written, with v the number of variables, as

$$\mathbf{y} = \sum_{j=0}^{v} \mathbf{x}_j \beta_j + \boldsymbol{\epsilon} \qquad (1.1)$$

with the elements of the vector ε assumed independently and normally distributed with zero mean and constant variance. (The main effect of variable j is usually defined to be twice the regression coefficient β_j.) If the above model were believed to be true, the parameters (including the error variance) could be efficiently estimated provided $(n-1)-\nu$ was large enough to provide desired degrees of freedom for estimating the variance, or if repeat runs were included for this purpose. A model of this form would be adequate when the response was roughly planar over the experimental region examined. On the other hand, allowance should be made for the possible inadequacy of the model (1.1). Suppose the true response function was much closer to a second-order model of the form

$$y = x_0\beta_0 + \sum_{j=1}^{\nu} x_j\beta_j + \sum_{i \le j}^{\nu} (x_i x_j)\beta_{ij} + \varepsilon. \qquad (1.2)$$

This would have the following implications. The estimate of the mean β_0 would be confounded with the pure quadratic coefficients β_{jj}. Estimates of the linear coefficients β_j may be confounded with interaction terms β_{ij}. The estimate of variance supplied by the $(n-1)-\nu$ unassigned columns may also be biased by real interaction effects.

To guard against the problems outlined above, one could take several approaches.

A second-order design could be employed which allowed estimation of all parameters of the model (1.2) (see, e.g., Box and Hunter, 1957). However, this greatly reduces the number of factors which could be studied in a given number of

experimental runs.

The inclusion of replicate runs in the two-level design would allow unbiased estimation of the variance. Lack of fit of the model (1.1) could be detected by the presence of large contrasts associated with the $(n-1)-v$ unassigned columns, and the design could be augmented to estimate the full second-order model, if necessary (Box and Wilson, 1951). However, the requirements of replicate runs again reduces the number of factors which could be studied in a given number of runs.

A third approach relies on a phenomenon of "effect sparsity" (Box and Meyer, 1985). The object of a screening experiment is to isolate important factors among a group of many candidates. If this is possible, then even if the true response was more closely approximated by the second-order model (1.2), many of the parameters would be negligible compared to the parameters associated with the important variables and the effect of noise. In this case an unreplicated two-level design will yield $n-1$ estimated effects, most of which will be inert and attributable to noise, the remainder of which will be active and too large to attribute to noise. As above, inadequacy of the first-order model could be detected by the presence of a large contrast associated with one of the $(n-1)-v$ unassigned columns.

This last approach, while combining the virtues of relatively low cost and relatively great information, does not always supply unambiguous results. Confounding of effects may lead to more than one plausible explanation of the data. However, a follow-up experiment to resolve ambiguities would usually involve fewer variables and many fewer runs than the original experiment, and the combined cost would be

less than the cost of a completely comprehensive experiment in all variables (Box, Hunter and Hunter, 1978).

1.3. Analysis of Fractional Factorials

Analysis of fractional factorial experiments has traditionally involved, primarily, identifying and estimating the active effects. In addition, estimating the error variance may also be of interest. The process of identifying the active effects has historically been divided into two stages (see, e.g., Box, Hunter and Hunter, 1978, Chapter 12). The first stage involves identifying the orthogonal contrasts $T_i = \mathbf{x}_i'\mathbf{y}/n$, $i=1,...n-1$, which are too large to be attributed to noise. These are called active contrasts. Under the second-order model (1.2) the expected value of T_i will be a linear combination of one or more of the coefficients β, sometimes called an alias string when involving more than one parameter. Under the hypothesis of effect sparsity, however, most of the contrasts will have expectation zero. A small proportion will have active terms in their alias string, and these will have non-zero expectation. The second stage of the analysis then involves determining which of the experimetal factors are associated with the active contrasts.

1.3.1. First stage analysis

Some of the techniques which have been employed to identify active contrasts are as follows.

Analysis of variance has been used to judge the reality of the contrasts (see Davies ed., 1954, p. 464). This method relies on comparison of the contrasts with

an independent measure of error variance. When an estimate of experimental error variance is available from relevant genuinely replicated runs from current or past experimentation, construction of the analysis of variance table is straightforward.

For unreplicated experiments, it has been customary to identify *a priori* certain contrasts, usually those which have only higher order interactions in their alias strings, whose magnitude could be attributed solely to random error. (In the case of quantitative factors, relative smoothness of the response surface would dictate that higher order interactions, which correspond to higher order derivatives, become successively smaller. This is reasonable as long as the ranges for the variables are chosen moderately. Likewise, for qualitative variables, the existence of higher order interactions implies a wide difference between levels of the variables, which should be avoided. Alternately, if the levels of qualitative variables must be chosen to be very dissimilar, separate experiments should be run for each level. In this way the frequency of large, high order interactions can be minimized, and contrasts which measure these interactions can be assumed to measure noise). These inert contrasts are then used to estimate error variance. This approach necessarily restricts the degree of fractionation to be used in the design, as several columns must be reserved to estimate effects supposedly known to be inert. Alternately when little is known about which effects are inert, the required contrasts may be difficult or impossible to identify. An even less satisfactory procedure for estimating the experimental error variance employs successive pooling of supposedly nonsignificant components in the analysis of variance table.

Daniel (1959) introduced the half-normal plot for judging the significance of orthogonal contrasts from a factorial experiment. In this method the $n-1$ ordered absolute contrasts $|T|_{(i)}$ are plotted against $\Phi^{-1}(1/2+(i-1/2)/2(n-1))$, where Φ is the standard normal distribution function. Under the completely null hypothesis of no active contrasts, these points should fall roughly along a straight line through the origin. Contrasts too large to be exlained by noise would appear as extreme points falling off the line. Later, Daniel (1976) pointed out that any information contained in the signs of the contrasts is obscured in the half-normal plot. A slight modification of Daniel's idea, the full-normal plot, i.e., plotting the signed ordered contrasts $T_{(i)}$ against $\Phi^{-1}((i-1/2)/(n-1))$, can be interpreted in the same way as the half-normal plot without losing the diagnostic information in the signs of the contrasts.

The advantages of normal probability plotting are that it requires neither replicated runs nor prior identification of inert contrasts and also allows for selection automatically. As with other graphical procedures, the normal plot may suggest further examination of the data. In particular, it can be used to detect model inadequacies (see Chapter 4).

Daniel also suggested how formal inference about which contrasts were significantly non-zero could be implemented through the normal plot. "Guardrails" of various Type I error rates are constructed by considering the null distributions of the ordered absolute contrasts. In a companion paper, Birnbaum (1959) discussed several methods for judging which contrasts measured non-zero effects, and showed that Daniel's procedure could be regarded as an approximation to the optimal statistic

when there was at most one significant contrast. In addition, Birnbaum stated that the optimal procedure for the case of more than one significant contrast was far too complicated for practical application, and concluded that Daniel's analysis was preferable for typical research applications. Zahn (1975) proposed some revisions to Daniel's procedure, including corrections to the critical values of the test.

Two other methods for analyzing unreplicated factorials were given by Wilk, Gnanadesikan and Freeny (1963) and Holms and Berrettoni (1969). Wilk, et al. suggested using maximum likelihood estimation of the variance, assuming that some number K of the original contrasts only measure error. The estimation is then based on the M ($<K$) smallest contrasts in order to avoid including contrasts measuring real effects in the estimate of σ, with suggested choice of M being $0.7K$. However, their estimate of σ was shown to be quite sensitive to the choice of K. Holms and Berrettoni proposed a method for the case when it is expected that a large proportion of the contrasts measure real effects. They considered the ordered absolute contrasts from smallest to largest, with each one in turn compared to those smaller than it. Critical values of the procedure, called "chain-pooling," were derived from work done by Cochran (1941).

1.3.2. Second stage analysis

Box and Hunter (1961) offered two guidelines for the process of associating factorial effects with active contrasts in the presence of confounding:

1. Main effects are more likely to occur than two-factor interactions, which are more likely than three-factor interactions, etc. That is, if a large contrast is

associated with more than one effect, the effect of lowest order is usually considered the most likely cause. This is especially true for continuous variables, when smoothness of the response surface dictates that higher-order effects, which correspond to higher-order derivatives, become successively smaller. In screening situations and other applications, it is common to ignore three-factor or higher order interactions.

2. Variables which have large main effects are more likely to have significant interactions among themselves or with other variables. For example, when a large contrast is associated with several two-factor interactions, the interactions involving variables with large main effects are considered more likely to be the cause.

The authors emphasize that these guidelines are to be employed to make tentative conclusions, subject to verification by subsequent experimentation or monitoring of the process after implementing changes. Exceptions to the rules appear, for example, when the design is located on a diagonal ridge of the response surface. This can occur when the process has been fine-tuned in the past one variable at a time, in the presence of compensating factors such as time and temperature of a chemical reaction. The experiment will then produce small main effects among the compensating factors, but a large two-factor interaction.

1.4. A Bayesian Approach

The assumptions that are made when analyzing factorials and fractional factorials can be modeled formally, and that is the basic premise of this thesis. Once the

assumptions are made explicit, Bayes' theorem provides a straightforward method of inference.

1.4.1. Identification of Active Contrasts

To model the assumption that a majority of the column contrasts are expected to be inert, it is assumed there is some prior probability α that each column is active, with α generally assumed to be less than 1/2. Let $a_{(c)}$ denote the event that a particular combination of c of the $n-1$ contrasts are active, the remainder inert. The prior probability of the event $a_{(c)}$ is

$$p(a_{(c)}) = \alpha^c (1-\alpha)^{n-1-c} .$$

After observing the data y from the experiment, the posterior probability of the event $a_{(c)}$ is

$$p(a_{(c)}|\mathbf{y}) = \frac{p(\mathbf{y}|a_{(c)})p(a_{(c)})}{\sum_{(i)} p(\mathbf{y}|a_{(i)})p(a_{(i)})} , \tag{1.3}$$

where the denominator is the summation over all possible combinations of active and inert columns, and $p(\mathbf{y}|a_{(c)})$ is the predictive density of the observations y given $a_{(c)}$. Of particular interest is the marginal posterior probability that column i is active, and this is given by

$$p_i = P[\text{column } i \text{ active}|\mathbf{y}] = \sum_{(c):i \text{ active}} p(a_{(c)}|\mathbf{y}) \tag{1.4}$$

Inference about which columns are active can be made from the probabilities $\{p_i\}$. The details of such an analysis are explored in Chapter 2.

1.4.2. Identification of Active Factors

Once active columns have been identified, it remains to identify which factors are responsible for the large contrasts. Alternately, in some situations there may only be interest in which factors are active, regardless of how their activity can be explained by main effects, interactions, etc. A modification of the Bayesian model introduced above is useful for this type of analysis.

Rather than contrasts being active with some prior probability α, it is assumed that factors will be active with probability α, with a suitable adjustment in the value of α. The notation $a_{(f)}$ would now refer to the event that a particular combination of f factors (including possible interactions) was active. The posterior probability of $a_{(f)}$ is then derived analagous to the expression for the posterior probability of $a_{(c)}$ given previously. The posterior probability of each factor being active is then

$$p_i^* = \sum_{(f):\, i \text{ active}} p(a_{(f)}|\mathbf{y}). \tag{1.5}$$

The details of the modified Bayesian model are given in Chapter 3, where it is demonstrated that the posterior probabilities $\{p_i^*\}$ take into account the confounding pattern of the design. Also included are simulation results for exploring the robustness of the posterior probabilities to the assumption of normally distributed errors.

1.5. Bad Values in Fractional Factorials

As noted by Daniel (1976) and others, the results of unreplicated fractional factorial experiments are sensitive to bad values among the observations. Daniel (1959) estimated that in his experience, the relative frequency of bad values in factorial

experiments was anywhere from .01 to .1, depending on the complexity of the experimental situation and on the experience of the experimenter. If, for example, each observation in a 16-run experiment had an independent probability of .05 of being incorrectly determined, then over half of such experiments would contain one or more bad values. Daniel also felt that quite often the presence of large higher-order interactions in factorial experiments was not due to highly curved response surfaces, but to erroneous observations which were not identified as such. Because of the saturated nature of unreplicated factorial experiments, bad observations can often be concealed by mistaken identification of some combination of active effects.

Full normal plotting of the observed contrasts (Daniel, 1959, 1976) has been a useful diagnostic tool for detecting bad values in unreplicated experiments, in addition to its use in identifying active contrasts. If a particular observation is biased by, say, a positive amount, those contrasts in which the observation enters positively are shifted to the right, and those contrasts in which the observation enters negatively are shifted to the left. This produces a "gap" among the inert contrasts of the normal plot which is the telltale sign of a bad observation. Similarly, the presence of multiple bad values can produce multiple gaps in the normal plot.

There is a wide literature on the general statistical issue of outliers. In their review article, Beckman and Cook (1983) list 229 references concerning the detection and accomodation or rejection of bad values. While some general regression diagnostics could conceivably be employed in the analysis of factorial experiments, Little (1983) for example showed that several of the common diagnostics consisted of a fac-

tor which measured the leverage of the doubtful points in the X space and a factor which measured the change in the residual sum of squares when suspected bad values were deleted. Because all design points have equal leverage in two-level factorials, for the case of one outlier the leverage factor would be the same for each observation, and in general one would not expect to have problems with extreme points in the X space from factorial experiments. Thus these diagnostics would reduce to functions of the change in the residual sum of squares when bad values are deleted, and some methods, described below, have been developed for factorial designs which are basically functions of the change in the residual sum of squares.

Daniel (1961) proposed a test for bad values based on the maximum residual after active contrasts have been identified. The observation corresponding to the maximum residual is identified as bad if the modulus of the residual is greater than a specified upper percentile of its null distribution. Stefansky (1972) derived revised critical values for Daniel's test.

John (1978) described a general method for detecting one or two bad values in a factorial experiment, based on work by Gentleman and Wilk (1975a,b), John and Prescott (1975a,b), and John and Draper (1978), which incorporates the reduction in the sum of squares when supposed bad values are deleted. The method is similar to one proposed by Goldsmith and Boddy (1973), and encompasses the test based on the maximum residual. It is described in more detail in Chapter 4.

The existing methodology for dealing with bad values in unreplicated factorials supposes that a fixed model has been identified. However, the possibility of bad

values may also be accomodated by "robustifying" the sampling distribution of y. Specifically, it is assumed the errors in the model (1.1) come from the scale-contaminated normal distribution denoted by

$$(1-\alpha_2)N(0,\sigma_\varepsilon^2)+\alpha_2 N(0,k_2^2\sigma_\varepsilon^2)$$

(Jeffreys, 1932; Dixon, 1951; Tukey, 1960; Box and Tiao, 1968). This provides for assessment of the contrasts while simultaneously allowing for the possibility of bad values, to be contrasted with the practice of checking for bad values after the active contrasts have been identified. Using the Bayesian approach above, the posterior probability that a contrast is active can be calculated while taking into account the possibility of bad values, and the posterior probability that an observation is "bad" can be calculated in light of the consideration that the identity of the active contrasts is not known. In this way questionable observations can be identified and investigated, and the sensitivity of the conclusions to the presence of possible bad values can be measured. Chapter 4 is devoted to details and discussion of this model.

The robustification of models in this way, while philosophically attractive, has been historically difficult to implement because it usually requires very extensive computing. However, as the speed and sophistication of computers has advanced, computationally intensive statistical analysis has become more feasible. At present the amount of computing time needed to analyze the above model allowing for bad values in full generality is not practical. Various computational shortcuts may be used, however, based on reasonable assumptions about the maximum possible number of bad values and active contrasts. It seems likely that future advances in computing

technology will reduce such limitations.

1.6. Posterior Probabilities For Model Selection

For the general problem of model discrimination, Atkinson (1978) presents three objections to the use of posterior probabilities:

1. When all of the candidate models fit badly, the best fitting of these will be chosen with high probability for n large enough.
2. When competing models have different numbers of parameters, the model with fewest parameters is favored in the absence of evidence in the data.
3. When models are nested and the simplest model is true, then all models are true and should receive equal weight. However, the simplest model will receive the highest posterior probability (by argument 2 above), and the remaining models will receive decreasing weights depending on the number of parameters.

The type of analysis proposed in this thesis could be classified as a model discrimination procedure. The objections listed above as they relate to the proposed analysis, are answered as follows:

1. Any statistical estimation procedure chooses the best-fitting model from a family of models according to some criterion. Usually this involves estimating parameters which index the family of models. However, there ought to be no implication in such a procedure that the best-fitting model among those considered will be adequate. Likewise, a model identified by a high posterior probability need not fit well. Diagnostic model-checking is an essential part of any statistical

analysis (see Box, 1980).

2. The validity of this objection for the general question of model discrimination will not be discussed here. Under the assumed condition of effect sparsity, the favoring of a model with fewer parameters, in the absence of evidence from the data, may be viewed as an advantage rather than a disadvantage.

3. When a model is "true", the statement that "all models in which the true model is nested are also true" is a matter of philosophy. One might argue that there is an inherent difference between the models

$$M_1: Y = \theta_1 X_1 + 0(X_2)$$

and

$$M_2: Y = \theta_1 X_1 + \theta_2 X_2.$$

In any event, the result that the simpler model receives higher posterior probability does not seem misleading to me.

CHAPTER 2

IDENTIFICATION OF ACTIVE CONTRASTS

2.1. Introduction

In a typical n-run unreplicated factorial or fractional factorial experiment, $n-1$ orthogonal contrasts can be computed. If one hypothesizes a model in which all experimental factors are active, then all or most of the contrasts will measure real effects. At the preliminary stage of experimentation, however, it is often felt that a large proportion of the factors will not be active, and therefore a large proportion of the contrasts will measure only noise. In this chapter the problem of identifying those contrasts which are too large to attribute to noise is considered, and a Bayesian solution is proposed.

Suppose that each contrast has some small probability α $(0<\alpha<1/2)$ of being active. Let $a_{(c)}$ be the event that a particular set of c of the $n-1$ contrasts are active, and $\tau_{(c)}$ be the vector of true effects corresponding to $a_{(c)}$. The prior probability of $a_{(c)}$ is $p^{(c)}=\alpha^c(1-\alpha)^{n-1-c}$. The conditional distribution of y given $a_{(c)}$ is

$$p(\mathbf{y}\mid\tau_{(c)},\sigma_\varepsilon,a_{(c)}) = (2\pi)^{-n/2}\sigma_\varepsilon^{-n}\times \exp\left\{\frac{-1}{2\sigma_\varepsilon^2}(\mathbf{y}-\mathbf{X}_{(c)}\tau_{(c)})'(\mathbf{y}-\mathbf{X}_{(c)}\tau_{(c)})\right\}. \tag{2.1}$$

where $\mathbf{X}_{(c)}$ are the columns of X corresponding to $a_{(c)}$. (It is assumed that the

overall mean τ_0 is always included in the model.) Suppose further that the prior distribution of the the elements of $\tau_{(c)}$ are independent and normal with zero mean and variance $\gamma^2\sigma^2$. If noninformative priors are specified for τ_0 and σ_ε (see Box and Tiao, 1973, Section 1.3), then it is assumed that the resulting analysis is well approximated by using

$$p(\tau_{(c)}|a_{(c)},\sigma_\varepsilon) \propto (\gamma\sigma_\varepsilon)^{-c}\exp\left\{\frac{-1}{2\sigma_\varepsilon^2}\tau_{(c)}'\Gamma_c\tau_{(c)}\right\} \tag{2.2}$$

$$p(\sigma_\varepsilon) \propto \frac{1}{\sigma_\varepsilon}$$

where the $c+1$ by $c+1$ matrix

$$\Gamma_c = \frac{1}{\gamma^2}\begin{bmatrix} 0 & 0' \\ 0 & \mathbf{I}_c \end{bmatrix}$$

and $\mathbf{I}_c$ is the c by c identity matrix.

To compute the posterior probability $p(a_{(c)}|\mathbf{y})$, define

$$h(\mathbf{y}|a_{(c)}) = \tag{2.3}$$

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} p(\mathbf{y}|\tau_{(c)},\sigma_\varepsilon,a_{(c)})p(\tau_{(c)}|\sigma_\varepsilon,a_{(c)})p(\sigma_\varepsilon)\,d\sigma_\varepsilon d\tau$$

the marginal predictive distribution of y given $a_{(c)}$. Then it is well known that the posterior probability of the event $a_{(c)}$ given y is

$$p(a_{(c)}|\mathbf{y}) = \frac{p^{(c)}h(\mathbf{y}|a_{(c)})}{\sum_{(c)} p^{(c)}h(\mathbf{y}|a_{(c)})}. \tag{2.4}$$

The probability $p(a_{(c)}|\mathbf{y})$ can be reexpressed as

$$p(a_{(c)}|\mathbf{y}) = C\ \frac{p^{(c)}h(\mathbf{y}|a_{(c)})}{p^{(0)}h(\mathbf{y}|a_{(0)})} \tag{2.5}$$

where

$$C = \frac{p^{(0)}h(\mathbf{y}|a_{(0)})}{\sum_{(c)} p^{(c)}h(\mathbf{y}|a_{(c)})}$$

is the constant which makes the probabilities sum to unity. The expression (2.5) gives $p(a_{(c)}|\mathbf{y})$ proportional to the posterior probability ratio that (i) $\tau_{(c)}$ are the active effects vs. (ii) there are no active effects. Specifically, this probability is written

$$p_{(c)} = p(a_{(c)}|\mathbf{y}) = C\left[\frac{\alpha}{1-\alpha}\right]^{c}\gamma^{-c}\times \tag{2.6}$$

$$\frac{|\mathbf{X}_{(0)}'\mathbf{X}_{(0)}|^{1/2}}{|\Gamma_c+\mathbf{X}_{(c)}'\mathbf{X}_{(c)}|^{1/2}}\left[\frac{S(\hat{\tau}_{(c)})+\hat{\tau}_{(c)}'\Gamma_c\hat{\tau}_{(c)}}{S(\tau_{(0)})}\right]^{-(n-1)/2}$$

where

$$\hat{\tau}_{(c)} = (\Gamma_c + \mathbf{X}_{(c)}'\mathbf{X}_{(c)})^{-1}\mathbf{X}_{(c)}'\mathbf{y} \tag{2.7}$$

and

$$S(\hat{\tau}_{(c)}) = (\mathbf{y}-\mathbf{X}_{(c)}\hat{\tau}_{(c)})'(\mathbf{y}-\mathbf{X}_{(c)}\hat{\tau}_{(c)})\,. \tag{2.8}$$

The probabilities $p(a_{(c)}|\mathbf{y})$ can be summed to give, for example,

$$p_i = P(\tau_i \text{ active}|\mathbf{y}) = \sum_{(c):\, i \text{ active}} p(a_{(c)}|\mathbf{y}). \tag{2.9}$$

The relative importance of column i may then be judged according to the size of p_i.

2.2. Computing the Posterior Probabilities

As described in the previous section, computing the probability p_i that contrast column i is active involves summation over the 2^{n-2} events $a_{(c)}$ in which i is supposed to be active. For n=16 this is 16,384 separate probability calculations. While this number can be handled rapidly enough on a high-speed computer, for n=32, say, the computation becomes unfeasible. However, there is an alternative Bayes factorization which allows the calculation of the probabilities $\{p_i\}$ without summing over all possible combinations of active contrasts. (The original factorization given in Chapter 1 remains useful for deriving approximations to the posterior distributions of τ and σ later in this chapter, as well as other derivations in subsequent chapters).

First of all, apply the one-to-one transformation $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}=\mathbf{T}$, and compute posterior distributions with respect to the new data T. The sampling distribution of T given $\{\tau,\sigma\}$ is

$$p(\mathbf{T}\,|\,\tau,\sigma) \;\alpha\; \sigma^{-n}\prod_{i=0}^{n-1}\exp\left\{\frac{-(T_i-\tau_i)^2}{2\sigma^2}\right\} \tag{2.10}$$

Thus each contast T_i is independent and normally distributed. The prior distribution of each expected contrast τ_i is

$$p(\tau_i\,|\,\sigma) = \alpha(2\pi)^{-1/2}(k^2-1)^{-1/2}\sigma^{-1}\exp\left\{-\frac{\tau_i{}^2}{(2(k^2-1)^{1/2}\sigma^2}\right\}+(1-\alpha)I[\tau_i=0] \tag{2.11}$$

where $k^2=n\gamma^2+1$ and

$$I[\tau_i=0] = \begin{cases} 0 \text{ if } \tau_i \neq 0 \\ 1 \text{ if } \tau_i = 0. \end{cases}$$

The prior distributions of $\log(\sigma)$ and τ_0 are uniform in the region where the likelihood is appreciable, and are aproximated by taking $p(\sigma,\tau_0)=1/\sigma$. Therefore the joint posterior distribution of $\{\tau,\sigma\}$ is

$$p(\tau,\sigma|T) \propto \sigma^{-n-1} \exp\left\{\frac{-(T_0-\tau_0)^2}{2\sigma^2}\right\} \times \tag{2.12}$$

$$\prod_{i=1}^{n-1}\left[(1-\alpha)\exp\left\{-\frac{T_i^2}{2\sigma^2}\right\} + \frac{\alpha}{(2\pi(k^2-1)^{1/2}\sigma} \exp\left\{\frac{-1}{2\sigma^2}\left[(T_i-\tau_i)^2 + \frac{\tau_i^2}{k^2-1}\right]\right\}\right].$$

Integrating τ out of this expression gives the marginal posterior distribution of σ,

$$p(\sigma|T) \propto \sigma^{-n} \prod_{i=1}^{n-1}\left[(1-\alpha)\exp\left\{-\frac{T_i^2}{2\sigma^2}\right\} + \frac{\alpha}{k}\exp\left\{-\frac{T_i^2}{2k^2\sigma^2}\right\}\right]. \tag{2.13}$$

The posterior probability $p_{i|\sigma}$ that τ_i is active, conditional on σ, is, by direct application of Bayes' theorem,

$$p_{i|\sigma} = \frac{\alpha p(T|\sigma,\tau_i \text{ active})}{\alpha p(T|\sigma,\tau_i \text{active}) + (1-\alpha)p(T|\sigma,\tau_i \text{notactive})}$$

$$= \frac{\frac{\alpha}{k}\exp\{-\frac{T_i^2}{2k^2\sigma^2}\}}{\frac{\alpha}{k}\exp\left\{-\frac{T_i^2}{2k^2\sigma^2}\right\} + (1-\alpha)\exp\left\{-\frac{T_i^2}{2\sigma^2}\right\}}. \tag{2.14}$$

Thus, conditional on σ, the posterior probability that τ_i is active depends only on the

data through the observed contrast T_i. The unconditional probability p_i can be computed from the simple expressions for $p_{i|\sigma}$ and $p(\sigma|T)$ by

$$p_i = \int_0^\infty p_{i|\sigma} p(\sigma|T)\, d\sigma. \tag{2.15}$$

This integral can then be computed to the desired degree of accuracy using numerical integration methods.

2.3. Prior Parameters k and α

2.3.1. Estimating a Range For k and α

In practice, α and k would be determined somewhat by the investigator's experience with the experimental material; the expected proportion of large effects represented by α and the expected size of such effects represented by k. The investigator may estimate these parameters from past similar experiments of his own or results reported in the scientific literature. To define a working range for α and k for the purposes of this thesis, the results of several unreplicated fractional factorials were examined. For each example, an estimate of α was obtained as the proportion of effects declared significant by the author(s) of that particular example, and k^2 was estimated by the ratio of the mean squared significant effect over the mean squared inert effect. These values are presented in Table 2.1. The estimated values of α range between .13 and .27 with an average of about 0.2. The estimated values of k range from 2.7 to 18 with an average of about 10. This gives an idea of plausible ranges for the two parameters.

Table 2.1 Estimated values of α and k from published examples of 16 and 32 run two-level designs taken from Box, Hunter and Hunter (1978), Davies ed. (1954), Daniel (1976), Bennett and Franklin (1954), Johnson and Leone (1964), and Taguchi and Wu (1980). In Daniel's example the analysis is conducted after making a log transformation in the response.

Example	n	α	k
BHH p. 398	16	.20	7.9
BHH p. 327	16	.27	13.9
BHH p. 378	32	.16	11.0
Davies p. 274	16	.13	2.7
Davies p. 462	16	.27	7.1
Daniel p. 71	16	.20	13.0
BF p. 557	16	.27	18.0
JL p. 183	32	.13	3.2
JL p. 196	16	.27	9.5
TW p. 69	16	.13	9.7
Average		.20	9.6

The possibility of bias introduced by restricting attention to published examples and by estimating α and k in this informal manner is recognized. However, it is shown later that the conclusions to be drawn from the analysis are usually insensitive to moderate changes in these parameters. In addition, convenient diagnostics exist which can detect those instances when the posterior probabilities are sensitive to such changes.

2.3.2. Two Examples

To illustrate an application, the following examples from Box, Hunter, and Hunter (1978), p. 398, and Davies ed. (1954), p. 274 were studied.

Example 2.1

The effects of 8 variables on the shrinkage y in an injection molding process were measured using a 2^{8-4} resolution IV design. The data are presented in Table 2.2a. A normal plot of the orthogonal contrasts, Figure 2.1a, revealed that there were two large main effects, due to holding pressure and booster pressure, and one other significant contrast aliased among four different two-factor interactions. (It was assumed that interactions between three or more factors were negligible).

Example 2.2

As described by the authors of Davies ed. (1954), this is a laboratory investigation in which interest centered on the effects of four factors on the yield of a product which was an isatin derivative. An unreplicated full 2^4 factorial experiment was run, and the data appear in Table 2.2b. Using analysis of variance with the

Table 2.2a Design matrix, observations, and observed contrasts for Example 2.1, a 2^{8-4} fractional factorial experiment, from Box, Hunter and Hunter (1978).

	factors								
run	1	2	3	4	5	6	7	8	y
1	-	-	-	+	+	+	-	+	14.0
2	+	-	-	-	-	+	+	+	16.8
3	-	+	-	-	+	-	+	+	15.0
4	+	+	-	+	-	-	-	+	15.4
5	-	-	+	+	-	-	+	+	27.6
6	+	-	+	-	+	-	-	+	24.0
7	-	+	+	-	-	+	-	+	27.4
8	+	+	+	+	+	+	+	+	22.6
9	+	+	+	-	-	-	+	-	22.3
10	-	+	+	+	+	-	-	-	17.1
11	+	-	+	+	-	+	-	-	21.5
12	-	-	+	-	+	+	+	-	17.5
13	+	+	-	-	+	+	-	-	15.9
14	-	+	-	+	-	+	+	-	21.9
15	+	-	-	+	+	-	+	-	16.7
16	-	-	-	-	-	-	-	-	20.3

column(effect)	observed contrast	column(effect)	observed contrast
0(mean)	19.75	8(8)	0.60
1(1)	-0.35	9(12+37+48+56)	-0.30
2(2)	-0.05	10(13+27+46+58)	0.45
3(3)	2.75	11(14+28+36+57)	-0.20
4(4)	-0.15	12(15+26+38+47)	2.30
5(5)	-1.90	13(16+25+34+78)	-0.15
6(6)	-0.05	14(17+23+68+45)	-0.10
7(7)	0.30	15(18+24+35+67)	-0.30

Table 2.2b Design matrix, observations, and observed contrasts for Example 2.2, a full 2^4 factorial experiment, from Davies ed. (1954).

	factors				
run	A	B	C	D	y(yield)
1	-	-	-	-	6.08
2	+	-	-	-	6.04
3	-	+	-	-	6.53
4	+	+	-	-	6.43
5	-	-	+	-	6.31
6	+	-	+	-	6.09
7	-	+	+	-	6.12
8	+	+	+	-	6.36
9	-	-	-	+	6.79
10	+	-	-	+	6.68
11	-	+	-	+	6.73
12	+	+	-	+	6.08
13	-	-	+	+	6.77
14	+	-	+	+	6.38
15	-	+	+	+	6.49
16	+	+	+	+	6.23

column(effect)	observed contrast	column(effect)	observed contrast
0(mean)	6.38	8(D)	.137
1(A)	-.096	9(AD)	-.082
2(B)	-.011	10(BD)	-.126
3(AB)	-.001	11(ABD)	-.051
4(C)	.038	12(CD)	-.013
5(AC)	-.017	13(ACD)	-.003
6(BC)	.033	14(BCD)	.062
7(ABC)	.075	15(ABCD)	.010

Figure 2.1a Full normal plot of observed contrasts, Example 2.1. The points are labelled by their column numbers.

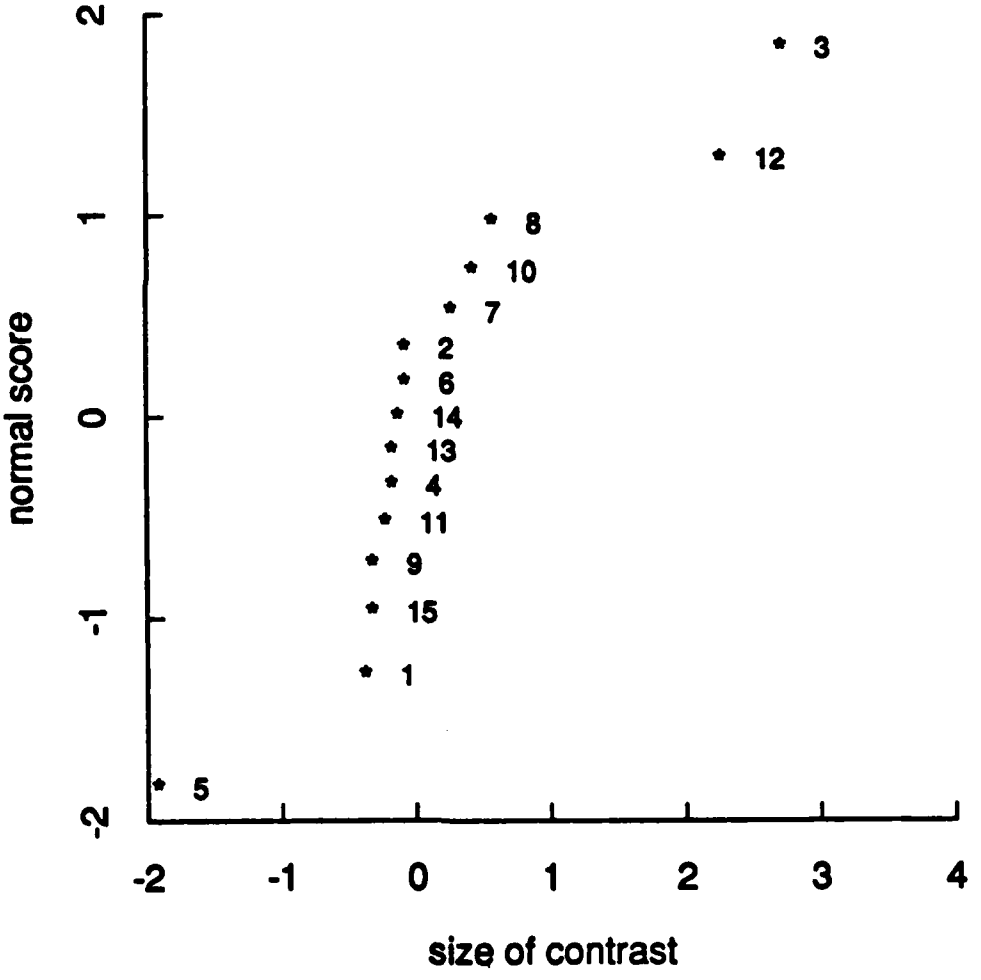
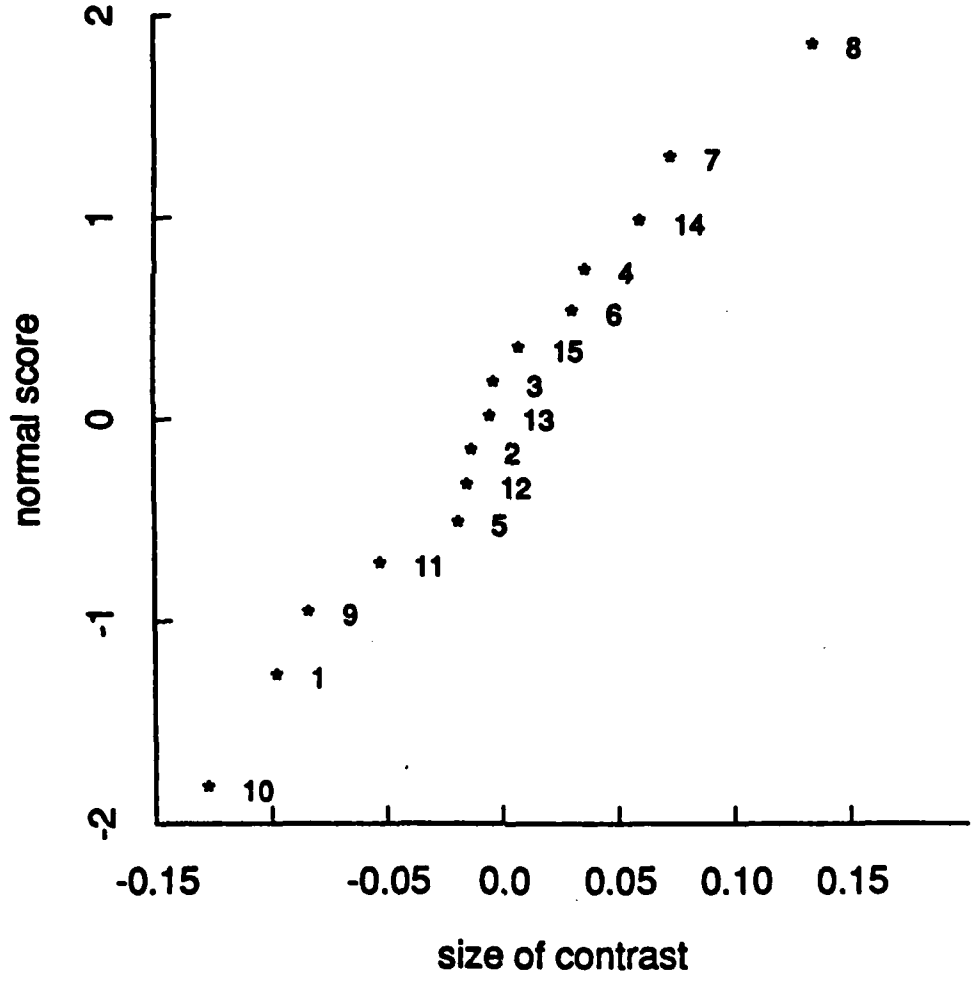


Figure 2.1b Full normal plot of observed contrasts, Example 2.2. The points are labeled by their column numbers.



three- and four-factor interactions serving as error, two contrasts corresponding to one main effect and one two-factor interaction were found to be marginally significant. The significance probabilities were close to .05; correcting for selection would have made them larger.. A normal plot of the orthogonal contrasts, Figure 2.1b, is consistent with these results, namely, there is little evidence for any active contrasts.

The posterior probabilities for the fifteen contrasts of Examples 2.1 and 2.2 were calculated and are presented in Figure 2.2. The solid vertical lines labelled 1 through 15 (corresponding to contrast columns 1 through 15) represent the probabilities calculated from (2.15) with the mean values α=0.2 and k=10. The boxes on each line indicate the range of the probabilities over all combinations of α=0.1, 0.2, 0.3 and k=5, 10, 15.

For Example 2.1, consider first the probabilities obtained with α=0.2 and k=10. There are three probabilities close to one, the rest more or less close to zero. This suggests a division into inert and active contrasts which agrees with the normal plot of Figure 2.1a. The changes in posterior probabilities obtained by varying α and k, indicated by the boxes in Figure 2.2a, are not such as to change conclusions about active and inert contrasts. Probabilities closest to zero or one tend to change very little, while the largest change occurs for the intermediate probability associated with column 8.

Example 2.2 was chosen to illustrate what can occur in a situation where the evidence for active effects is much weaker. The probabilities obtained by setting α=0.2

Figure 2.2a Posterior probabilities $\{p_i\}$ that columns are active, Example 2.1. Solid vertical lines are the values for α=0.2 and k=10; boxes indicate the range of values over α=0.1, 0.2, 0.3 and k=5, 10, 15.

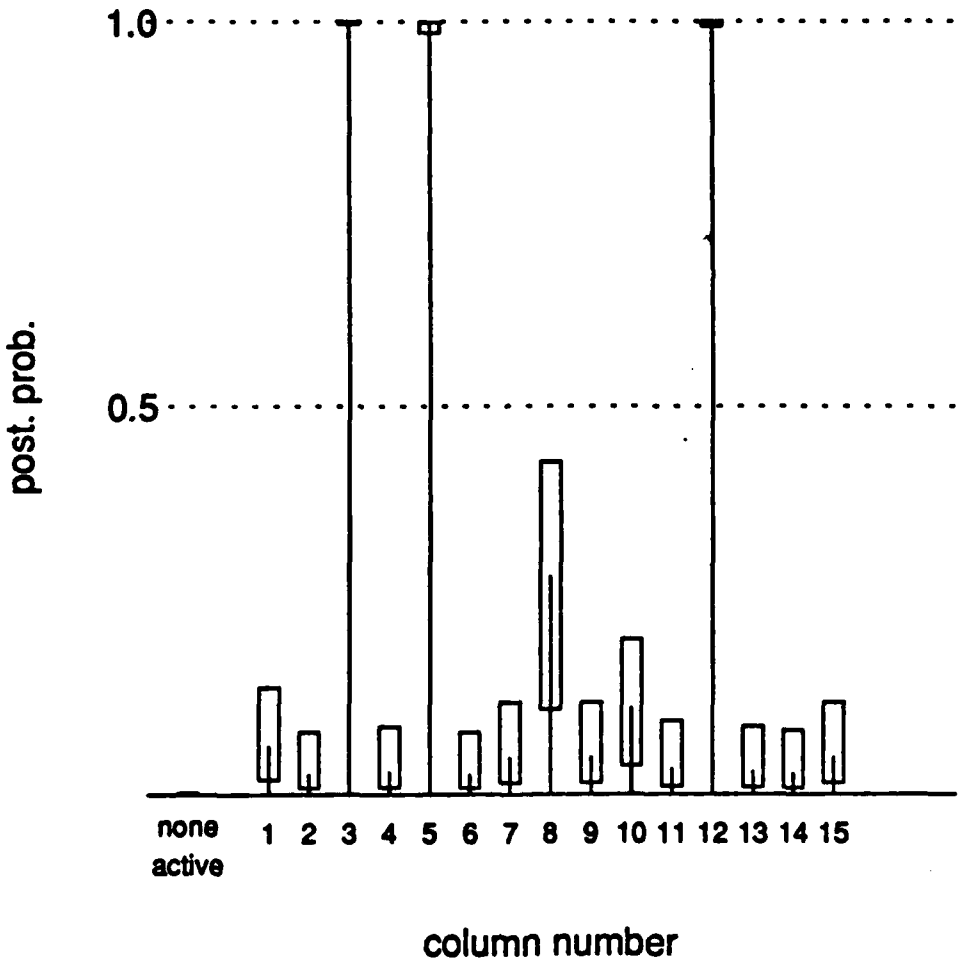
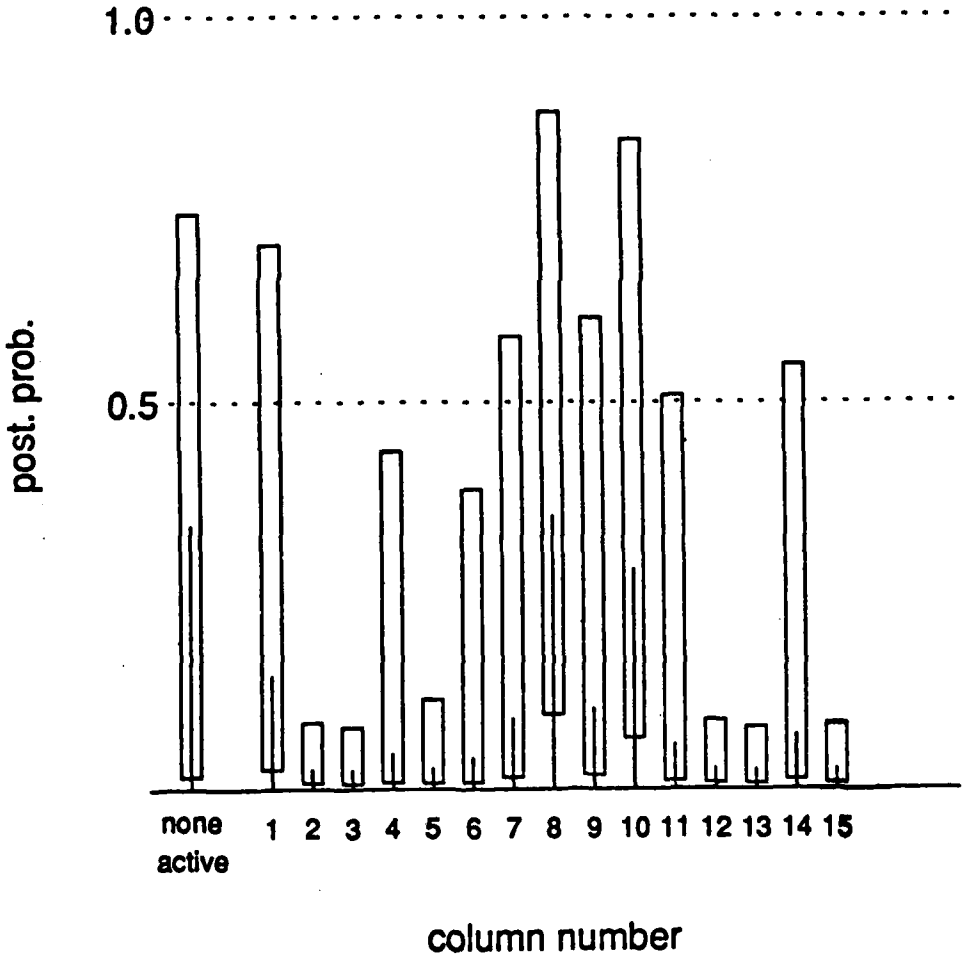


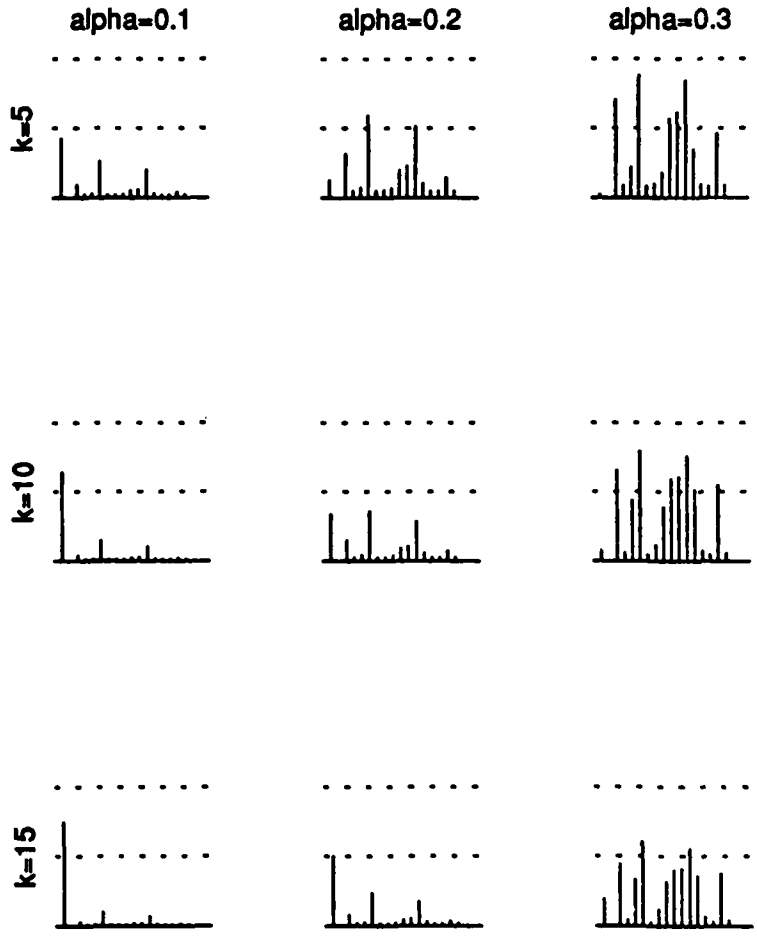
Figure 2.2b Posterior probabilities $\{p_i\}$ that columns are active, Example 2.2. Solid vertical lines are the values for α=0.2 and k=10; boxes indicate the range of values over α=0.1, 0.2, 0.3 and k=5, 10, 15.



and $k=10$ indicate that there is little evidence for any of the contrasts being active, or perhaps a weak suggestion of activity for columns 8 and 10, agreeing with analysis of variance results and the normal plot of Figure 2.1b. However, there is a much greater sensitivity to variation of α and k, as indicated by the boxes in Figure 2.2b, than for Example 2.1. To allow more detailed study the posterior probabilities for Example 2.2 are plotted individually in Figure 2.3 for each combination of α and k. In particular note that for $\alpha=0.3$ and $k=10$ the posterior probabilities for columns 1, 7, 8, 9, 10, 11 and 14 are all greater than 1/2. The situation can be further understood by reexamining the normal plot for this example in Figure 2.1b. At first glance it appears that all of the contrasts fall along a straight line more or less. On the other hand one could draw a line through the middle eight contrasts or so and declare the remaining seven to be active. This second interpretation would agree with a prior belief that there was a larger proportion of active contrasts, corresponding to the Bayesian analysis with a larger value of α. In light of this, it would be impossible to make reliable inferences about active and inert contrasts, because the conclusions change quite dramatically under differing plausible model assumptions.

To more closely follow the intent of the original authors of Example 2.2, the assumption that certain columns, corresponding to higher-order interactions, are inert can be incorporated easily into the Bayesian analysis by assigning a prior probability of zero to those particular columns. When this is done the posterior probabilities of the remaining columns are very close to those obtained in the above analysis with $\alpha=0.2$ for all columns.

Figure 2.3 Individual posterior probability plots for all combinations of α=0.1, 0.2, 0.3 and k=5, 10, 15, Example 2.2.



It should be recognized that the troublesome behavior exhibited in Example 2.2 is not due to a shortcoming of the procedure proposed here but rather to a lack of information in the data. This illustrates a point made by Barnard (1980), that there exist robust and non-robust data samples. For some sets of data, analyses undertaken over a plausible range of assumptions lead to essentially the same conclusions, while for others the conclusions are quite sensitive to changing assumptions. Thus with the robust data of Example 2.1 variation of α and k produces little change in the conclusions, while for the non-robust data of Example 2.2 the conclusions are quite sensitive to changing α and k.

2.3.3. Derivatives of the Posterior Probabilities

As illustrated by Example 2.2, there will be occasions when the probabilities $\{p_i\}$ will be sensitive to the choice of α and k. The partial derivatives of p_i with respect to α and k can be computed to measure the extent to which the probabilities are sensitive to the particular choice of these parameters.

First define the following quantities:

$$p_{ij} = P[\tau_i,\tau_j \text{ active} \mid \mathbf{T}] = \begin{cases} \int_0^\infty p_{i|\sigma} p_{j|\sigma} p(\sigma \mid \mathbf{T})\, d\sigma & (i \neq j) \\ p_i & (i = j) \end{cases} \tag{2.16}$$

$$Q_j = \left[\frac{T_j^2}{\sigma^2} - k^2\right]. \tag{2.17}$$

Then the partial derivatives of p_i with respect to α and k are given by

$$\frac{\partial p_i}{\partial \alpha} = \frac{1}{\alpha(1-\alpha)} \left[\sum_{j=1}^{n-1} (p_{ij} - p_i p_j) \right] \tag{2.18}$$

$$\frac{\partial p_i}{\partial k} = \frac{1}{k^3} \left[\int_0^\infty (p_{i|\sigma} - p_i)\left(\sum_{j=1}^{n-1} Q_{\ j} p_{j|\sigma} \right) p(\sigma|T)\, d\sigma \right.$$

$$\left. + \int_0^\infty p_{i|\sigma}(1-p_{i|\sigma}) Q_{\ i} p(\sigma|T)\, d\sigma \right]. \tag{2.19}$$

(Recall that the quantities p_i, p_{ij}, $p_{i|\sigma}$ and $p(\sigma|T)$ in the above expressions all also depend on α and k.)

The use of derivatives to measure sensitivity relies somewhat on a low degree of curvature in p_i with respect to α and k. In Figure 2.4 the posterior probabilities for Examples 2.1 and 2.2 are plotted first against α for fixed $k=10$, and then against k for fixed $\alpha=0.2$. In all figures the relative curvature is not so extreme that the partial derivatives would not give a good measure of change.

Consider first the derivative with respect to α, which is proportional to

$$\sum_{j=1}^{n-1} (p_{ij} - p_i p_j).$$

For columns for which p_i is close to one, p_{ij} will be approximately equal to p_j, all terms in the summation will be negligible, and the derivative will thus be close to zero. Similarly, if p_i is close to zero, p_{ij} and $p_i p_j$ will both be small and roughly equal, and again the derivative will be close to zero. For moderate p_i (in the neigh-

Figure 2.4a Continuous plot of posterior probabilities $\{p_i\}$ versus α, k fixed at 10, Example 2.1. Curves are labeled on the right by their column numbers.

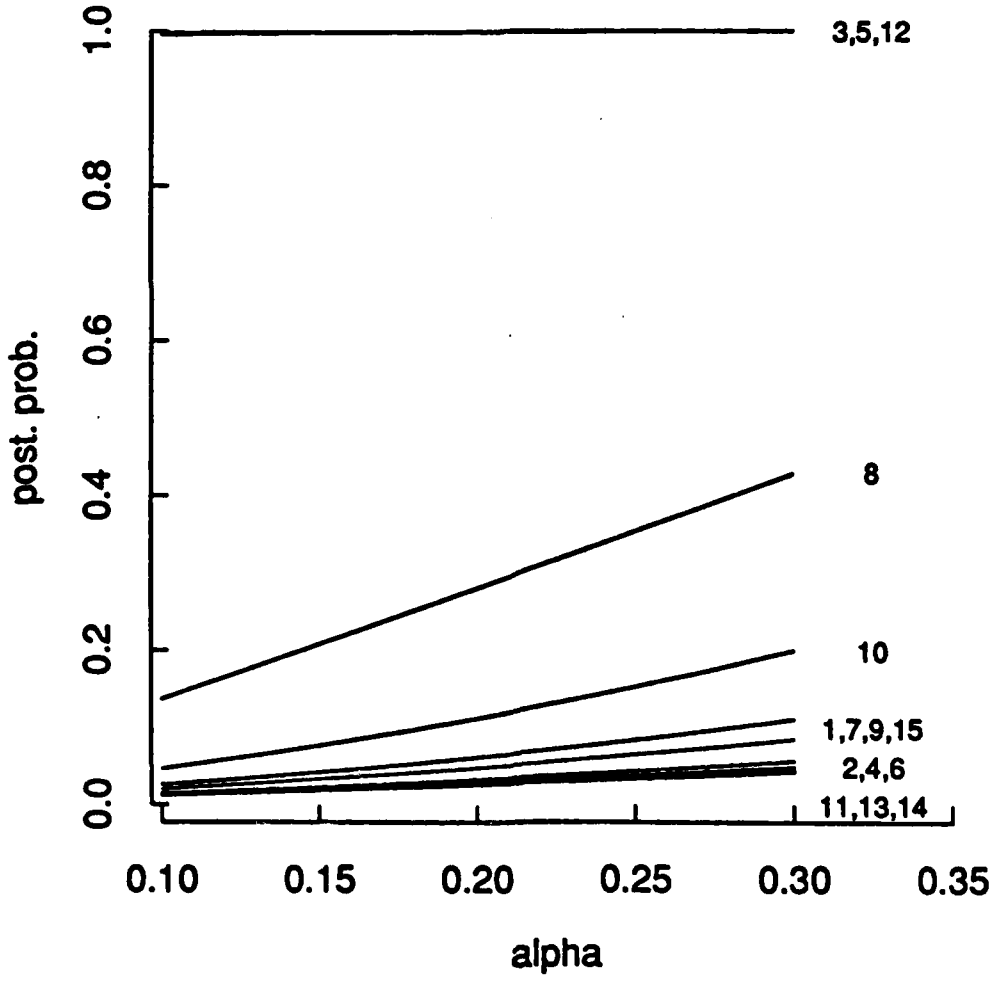


Figure 2.4b Continuous plot of posterior probabilities $\{p_i\}$ versus k, α fixed at 0.2, Example 2.1. Curves are labeled on the right by their column numbers.

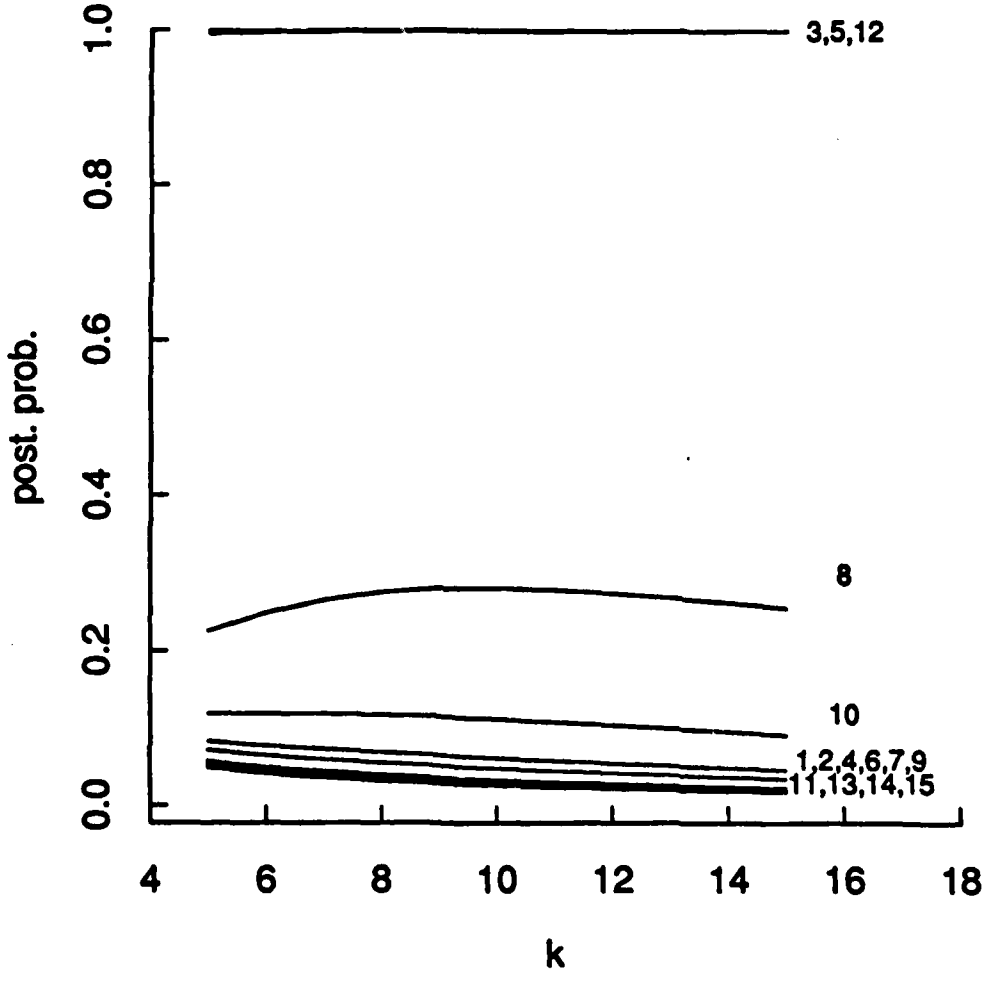


Figure 2.4c Continuous plot of posterior probabilities $\{p_i\}$ versus α, k fixed at 10, Example 2.2. Curves are labeled on the right by their column numbers.

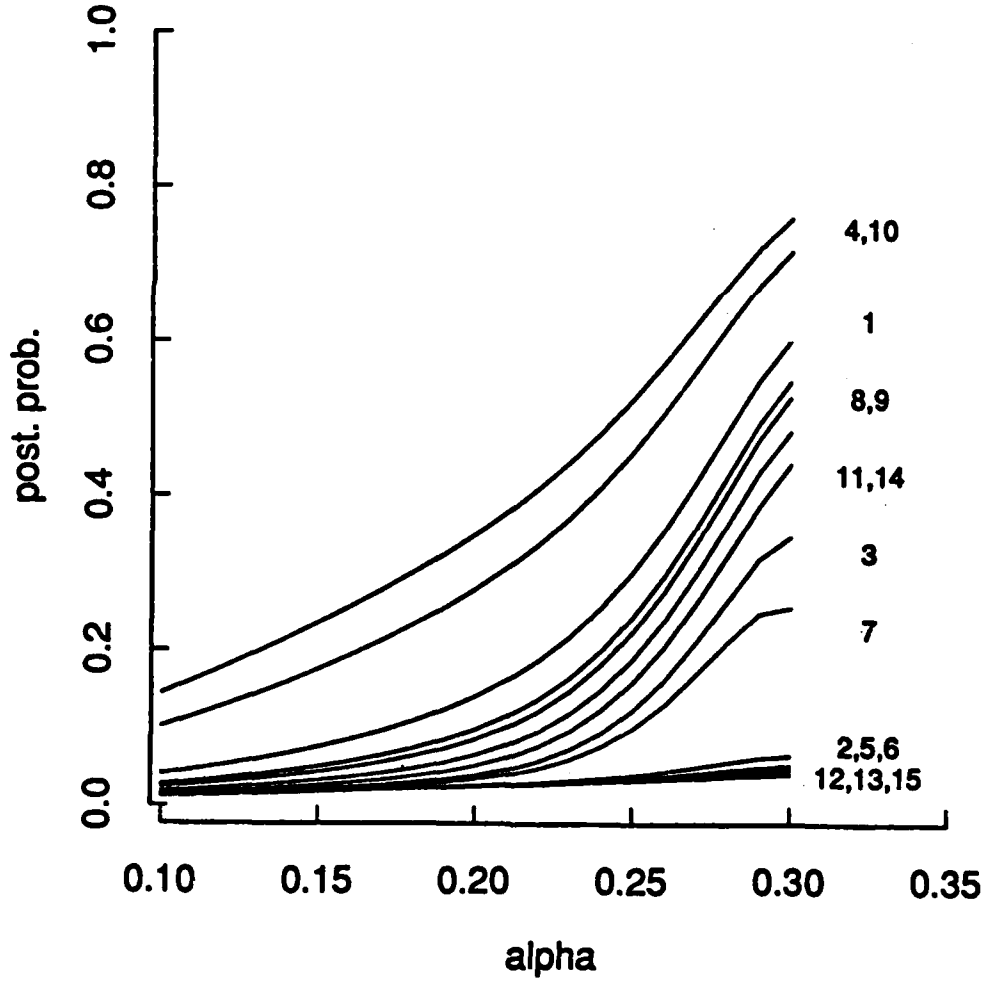
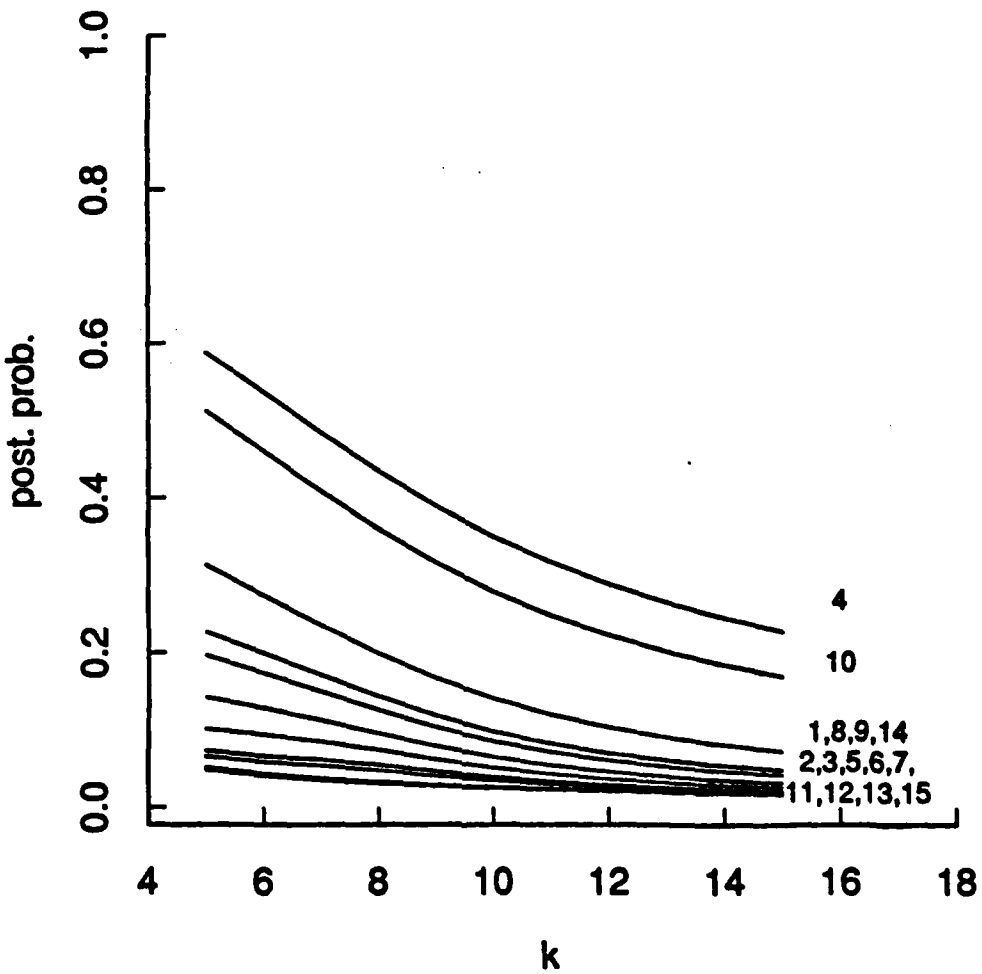


Figure 2.4d Continuous plot of posterior probabilities $\{p_i\}$ versus k, α fixed at 0.2, Example 2.2. Curves are labeled on the right by their column numbers.



borhood of 1/2), it will be helpful to define

$$p_{i|j} = P[\tau_i \text{active} \,|\, T, \tau_j \text{active}] \tag{2.20}$$

$$= \frac{p_{ij}}{p_j}.$$

Then the summand in (2.18) can be written

$$p_{ij} - p_i p_j = (p_{i|j} - p_i) p_j.$$

Now consider $p_{i|j}$ for columns j with very large contrast T_j. The information that τ_j is active will have very little effect on the probability that τ_i is active (this is generally true when conditioning on an event of probability close to one), and $p_{i|j}$ will be close to p_i. However, as the value of T_j approaches T_i, the information that τ_j is active gives more strength to the possibility that τ_i is active, and $p_{i|j}$ will become larger than p_i. Thus, at least for moderate p_i, the nonnegligible terms in the sum in 2.9 will be positive. Furthermore, the contribution from the term $p_i - p_i^2$ alone will be fairly large. (For p_i between α and $(1-\alpha)$, this term is greater than one; a derivative greater than one implies that any local change in α results in an even larger change in p_i). There will also be significant contributions from other terms for which p_j is close to 1/2, so that the value of the derivative will depend to a large extent on the total number of probabilities p_j not close to zero or one.

The expression for the derivative with respect to k is more complicated but it is still possible to interpret it. It should be noted first that while the probabilities p_i are more or less monotonic in α, this is not true for k. For example, in Figure 2.4c, it is seen that the some of the probabilities reach a maximum in the area of $k=10$, although

the plot exhibits very little curvature. Thus the derivative with respect to k could conceivably be close to zero even when the probabilities are actually sensitive to the choice of k, if k has been chosen near this maximum. However, the derivative can be interpreted as a measure of how well the chosen value of k fits the data. Consider the factor

$$\sum_{j=1}^{n-1} Q_j p_{j|\sigma} = \sum_{j=1}^{n} \left[\frac{T_j^2}{\sigma^2} - k^2 \right] p_{j|\sigma}$$

in the first integral in the expression (2.19) for the derviative of p_i with respect to k. It will be small if Q_j is small when $p_{j|\sigma}$ is close to one. For example, suppose the observed contrasts were easily partitioned into an inert group and an active group. If k^2 was chosen to be the ratio of the mean squared active contrast over the mean squared inert contrast, the nonnegligible terms in the sum would cancel. Thus a small value of this derivative can indicate either insensitivity to the choice of k or, lacking that, that k has been chosen to fit the data well in the sense described above. In either case the derivative is a useful diagnostic for this model.

While the first integral in (2.19) gives a rough estimate of how well k fits the data, the second integral is more a function of the individual contrast T_i. The factor $p_{i|\sigma}(1-p_{i|\sigma})$ achieves a maximum at $p_{i|\sigma}=1/2$, showing that the moderate p_i are also most sensitive to changes in k as measured by the partial derivative.

The derivatives of p_i with respect to α and k for Examples 2.1 and 2.2 are given in Table 2.3. The derivative with respect to k in the table is multiplied by 50, so that the two derivatives are roughly comparable. Certainly absolute values for these

Table 2.3 Posterior probabilities and derivatives with respect to α and k, Examples 2.1 and 2.2.

Example 2.1

column	observed contrast	posterior prob.	$dp/d\alpha$	$50(dp/dk)$
1	-0.35	0.0455	0.4163	-0.1783
2	-0.05	0.0167	0.1517	-0.1203
3	2.75	0.9998	0.0025	-0.0004
4	-0.15	0.0195	0.1784	-0.1311
5	-1.90	0.9987	0.0124	0.0021
6	-0.05	0.0167	0.1517	-0.1203
7	0.30	0.0342	0.3156	-0.1666
8	0.60	0.2548	1.4628	-0.047
9	-0.30	0.0342	0.3156	-0.1666
10	0.45	0.0910	0.7605	-0.1738
11	-0.20	0.0225	0.2062	-0.1408
12	2.30	0.9995	0.0050	-0.0002
13	-0.15	0.0195	0.1784	-0.1311
14	-0.10	0.0177	0.1611	-0.1243
15	-0.30	0.0342	0.3156	-0.1666

Example 2.2

column	observed contrast	posterior prob.	$dp/d\alpha$	$50(dp/dk)$
1	-.096	0.1518	1.8788	-1.1367
2	-.011	0.0293	0.1828	-0.1907
3	-.001	0.0407	0.5045	-0.2525
4	.038	0.3683	2.7261	-1.845
5	-.017	0.0275	0.1532	-0.1597
6	.033	0.0285	0.1820	-0.1666
7	.075	0.0339	0.3524	-0.1932
8	.137	0.0887	1.3115	-0.6493
9	-.082	0.1063	1.4543	-0.7989
10	-.126	0.3076	2.6355	-1.7825
11	-.051	0.0545	0.8433	-0.3882
12	-.013	0.0288	0.1855	-0.1805
13	-.003	0.0287	0.1731	-0.1861
14	.062	0.0709	1.0898	-0.5362
15	.010	0.0270	0.1690	-0.1615

derivatives far above one indicate acute sensitivity to the particular choice of parameter, while values far below one indicate insensitivity. The values in Table 2.3 agree with the above arguments and also with Figure 2.2.

2.4. Posterior distributions

Information about model parameters is summarized in their posterior distribution according to the Bayesian paradigm. Parametric inference will consist of interpreting the posterior distribution, for example, by calculating point parameter estimates or constructing Bayesian confidence regions, also called highest posterior density (H.P.D.) regions (Box and Tiao, 1973, p. 110). For analysis of the effect sparsity model, identifying and estimating the active contrasts is of primary interest. In some cases there will be interest in estimating the variance σ^2. These are achieved through calculation of the marginal posterior distributions of τ and σ respectively.

The derivations are given in terms of the original Bayes factorization of Section 2.1, which has greater intuitive appeal. Actual computations would be done in terms of the alternative factorization of Section 2.2. Relevant details are given following the derivations.

2.4.1. Joint posterior of $\{\tau,\sigma\}$

Straightforward application of Bayes' theorem would ordinarily result in the following expression for the joint posterior distribution of $\{\tau,\sigma\}$:

$$p(\tau,\sigma|y) \;\alpha\; p(y|\tau,\sigma)p(\tau|\sigma)p(\sigma). \tag{2.21}$$

However, a simpler factorization is obtained by conditioning on the events $a_{(c)}$ and marginalizing as follows:

$$p(\tau,\sigma|y) = \sum_{(c)} p(\tau,\sigma|y,a_{(c)})p(a_{(c)}|y). \tag{2.22}$$

The probabilities $p(a_{(c)}|y)$ have been derived previously (2.9, 2.15). The conditional posterior distribution $p(\tau,\sigma|y,a_{(c)})$ is given by:

$$p(\tau,\sigma|y,a_{(c)}) \;\alpha\; p(y|\tau,\sigma,a_{(c)})p(\tau|\sigma,a_{(c)})p(\sigma|a_{(c)})$$

$$= \sigma^{-n}\exp\left\{\frac{-1}{2\sigma^2}(y-X_{(c)}\tau_{(c)})'(y-X_{(c)}\tau_{(c)})\right\}\times$$

$$(\gamma\sigma)^{-c}\exp\left\{\frac{-1}{2\sigma^2}\tau_{(c)}'\Gamma_c\tau_{(c)}\right\}\times\sigma^{-1}$$

$$= \gamma^{-c}\sigma^{-n-c-1}\exp\left\{\frac{-1}{2\sigma^2}[S(\tau_{(c)})+\tau_{(c)}'\Gamma_c\tau_{(c)}]\right\}, \tag{2.23}$$

where

$$S(\tau_{(c)}) = (y-X_{(c)}\tau_{(c)})'(y-X_{(c)}\tau_{(c)}) \tag{2.24}$$

and recall Γ_c is defined by

$$\Gamma_c = \frac{1}{\gamma^2}\begin{bmatrix}0 & 0'\\ 0 & I_c\end{bmatrix}. \tag{2.25}$$

2.4.2. Marginal Posterior Distribution of τ

The marginal posterior distribution of τ is obtained from the joint posterior distribution of $\{\tau,\sigma\}$ by

$$p(\tau|y) = \int_0^\infty p(\tau,\sigma|y)\,d\sigma$$

$$= \int_0^\infty \sum_{(c)} p(\tau_{(c)},\sigma|y,a_{(c)})\,p(a_{(c)}|y)\,d\sigma$$

$$= \sum_{(c)} \left[\int_0^\infty p(\tau_{(c)},\sigma|by,a_{(c)})\,d\sigma\right] p(a_{(c)}|y)$$

$$= \sum_{(c)} p(\tau_{(c)}|y,a_{(c)})\,p(a_{(c)}|y). \tag{2.26}$$

Thus the marginal posterior of τ is a weighted sum of conditional posterior distributions given $a_{(c)}$. To find the conditional posterior of $\tau_{(c)}$,

$$p(\tau_{(c)}|y,a_{(c)}) = \int_0^\infty \gamma^{-c}\sigma^{-n-c-1}\exp\left\{\frac{-1}{2\sigma^2}[S(\tau_{(c)})+\tau_{(c)}'\Gamma_c\tau_{(c)}]\right\}d\sigma$$

$$\alpha\ \gamma^{-r}[S(\tau_{(c)})+\tau_{(c)}'\Gamma_c\tau_{(c)}]^{-(n+c)/2}. \tag{2.27}$$

It can be shown that

$$S(\tau_{(c)})+\tau_{(c)}'\Gamma_c\tau_{(c)} = (\tau_{(c)}-\hat{\tau}_{(c)})'(\Gamma_c+\mathbf{X}_{(c)}'\mathbf{X}_{(c)})(\tau_{(c)}-\hat{\tau}_{(c)})$$
$$+ S(\hat{\tau}_{(c)})+\hat{\tau}_{(c)}'\Gamma_c\hat{\tau}_{(c)} \tag{2.28}$$

where

$$\hat{\tau}_{(c)} = (\Gamma_c+\mathbf{X}_{(c)}'\mathbf{X}_{(c)})^{-1}\mathbf{X}_{(c)}'y. \tag{2.29}$$

Thus, after normalizing, the conditional posterior distribution of $\tau_{(c)}$ is

$$p(\tau_{(c)}|y,a_{(c)}) = \frac{\Gamma((n+c-1)/2)}{\Gamma((n-1)/2)} \frac{|\Gamma_c+X_{(c)}'X_{(c)}|^{1/2}}{(\pi(n-1)s^2_{(c)})^{(c+1)/2}} \times \tag{2.30}$$

$$\left[1+\frac{(\tau_{(c)}-\hat{\tau}_{(c)})'(\Gamma_c+X_{(c)}'X_{(c)})(\tau_{(c)}-\hat{\tau}_{(c)})}{(n-1)s^2_{(c)}}\right]^{-(n+c)/2},$$

where

$$s^2_{(c)} = \frac{(S(\hat{\tau}_{(c)})+\hat{\tau}_{(c)}'\Gamma_c\hat{\tau}_{(c)})}{n-1}. \tag{2.31}$$

This is a $(c+1)$-dimensional multivariate t distribution with $n-1$ degrees of freedom, mean vector $\hat{\tau}_{(c)}$, and dispersion matrix

$$(\Gamma_c+X_{(c)}'X_{(c)})s^2_{(c)}.$$

The complete marginal posterior distribution of a single τ_i is then

$$p(\tau_i|y) = \sum_{(c)} p(\tau_i|y,a_{(c)})p(a_{(c)}|y)$$

$$= (1-p_i)I[\tau_i=0] + \sum_{(c):i \text{ active}} p(\tau_i|y,a_{(c)})p(a_{(c)}|y), \tag{2.32}$$

a mixture of a discrete and continuous distribution. The continuous part is a weighted sum of t densities with $n-1$ degrees of freedom, common mean $\hat{\tau}_i$, different scale factors $s_{(c)}$, and weights $p(a_{(c)}|y)$. The discrete part has mass $1-p_i$ at zero.

It was shown previously how the relative activity of each contrast column can be measured by the posterior probability p_i. This accounts for the first term of (2.32), the mass $1-p_i$ at zero. In many instances parameter estimates and confidence bounds for

the supposed active τ_i will be desired, that is, those for which p_i is close to one. These can be obtained from the conditional posterior distribution of τ_i given it is active,

$$p(\tau_i|\mathbf{y},\tau_i \text{ active}) = \frac{1}{p_i} \sum_{(c):i \text{ active}} p(\tau_i|\mathbf{y},a_{(c)})p(a_{(c)}|\mathbf{y}), \tag{2.33}$$

the second term of (2.32) normalized to integrate to unity. An accepted point estimate of τ_i is the conditional posterior mean which can be written

$$\hat{\tau}_i = \left[n + \frac{1}{\gamma^2}\right]^{-1} \mathbf{x}_i'\mathbf{y} = \phi T_i, \tag{2.34}$$

which would be very close to the usual column contrast T_i. Confidence bounds of probability $1-\delta$ are obtained by specifying b_δ so that

$$P[\hat{\tau}_i - b_\delta < \tau_i < \hat{\tau}_i + b_\delta|\mathbf{y}] = 1-\delta. \tag{2.35}$$

In general, calculation of the exact posterior distribution and corresponding confidence bounds for τ_i would be very cumbersome. In practical application of fractional factorials, some sort of approximation would be needed. In particular, obtaining confidence limits by reference to a standard table with the aid of a convenient summary statistic would be much more appealing. It will be shown that in many cases the posterior distribution of an expected contrast τ_i can be well approximated by a single t distribution with $n-1$ degrees of freedom, mean $\hat{\tau}_i$, and scale factor

$$s_i^2 = \sum_{(c):i \text{ active}} s^2_{(c)} p(a_{(c)}|\mathbf{y}). \tag{2.36}$$

For illustration, I have taken the largest contrast in Example 2.1, and its "exact" conditional posterior density was computed by evaluating the weighted sum of t densities only over those $a_{(c)}$ with posterior probability larger then 0.0001, with α=0.2, γ=2.5. This accounted for over 99% of the total probability. The approximate density was computed as a single t density with scale factor s^2 defined above. The exact distribution is plotted as a solid curve in Figure 2.5a, and the approximate distribution is plotted as a dashed curve. For this example approximation by a single t distribution is very good.

For Example 2.2, none of the contrasts had large posterior probability for α=0.2 and γ=2.5. However, for α=0.3 and γ=1.25, contrast 8 has posterior probability .85, and the exact and approximate conditional posterior densities for this contrast are shown in Figure 2.5b. The approximation for this example is less accurate than for Example I.

The accuracy of the approximation depends upon the extent to which the t density is linear in the parameter s^2 over the range of $s^2_{(c)}$ in the weighted sum. To see this, write the weighted sum of densities generically as

$$p(\tau|y) = \sum_{(c)} p_{(c)}\, t(\tau|s_i^2) \tag{2.37}$$

where $t(\tau|s^2_{(c)})$ is the t density with scale parameter $s^2_{(c)}$. Then the approximate density is just

$$\hat{p}(\tau|y) = t(\tau|\hat{s}^2)\,. \tag{2.38}$$

If the function $t(\tau|s^2)$ was linear in s^2, or if the scale factor $s^2_{(c)}$ did not vary over

Figure 2.5a Exact and approximate posterior densities for largest expected contrast, Example 2.1, α=0.2, γ=2.5. The exact (solid curve) is obtained by direct calculation and the approximate (dashed curve) is a single t density.

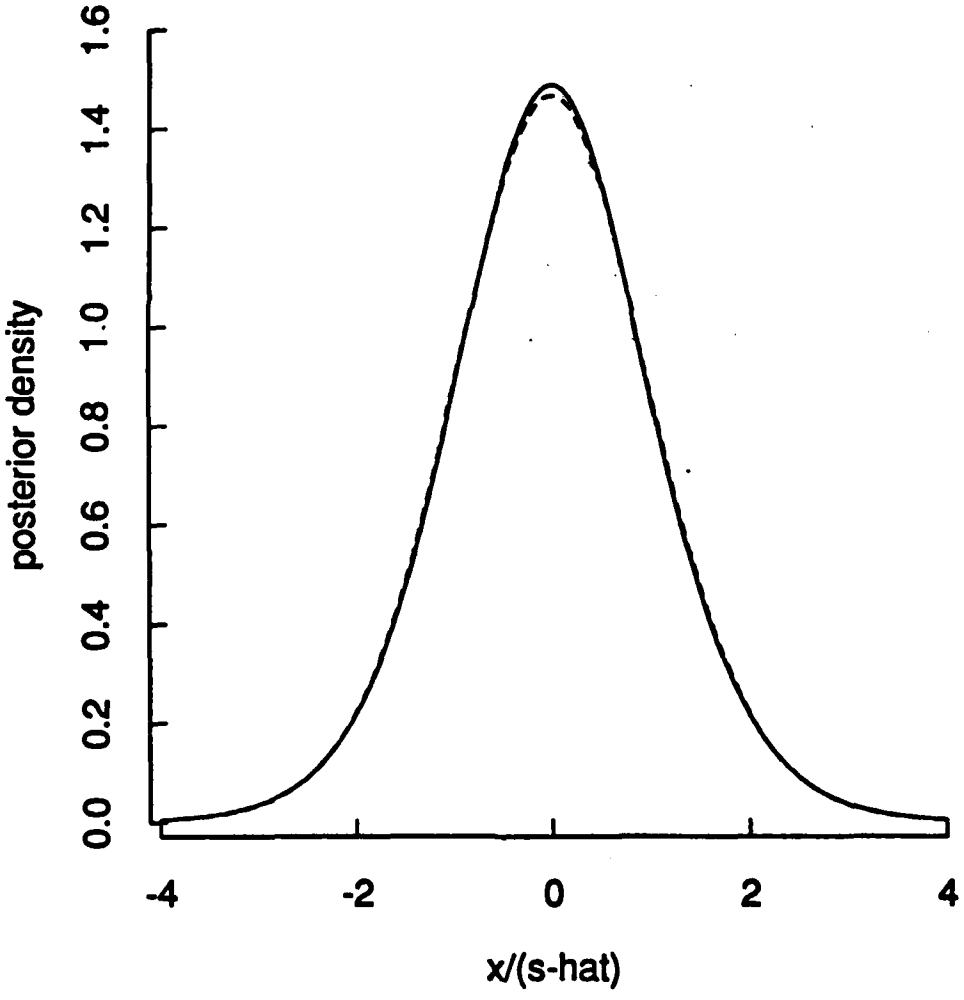
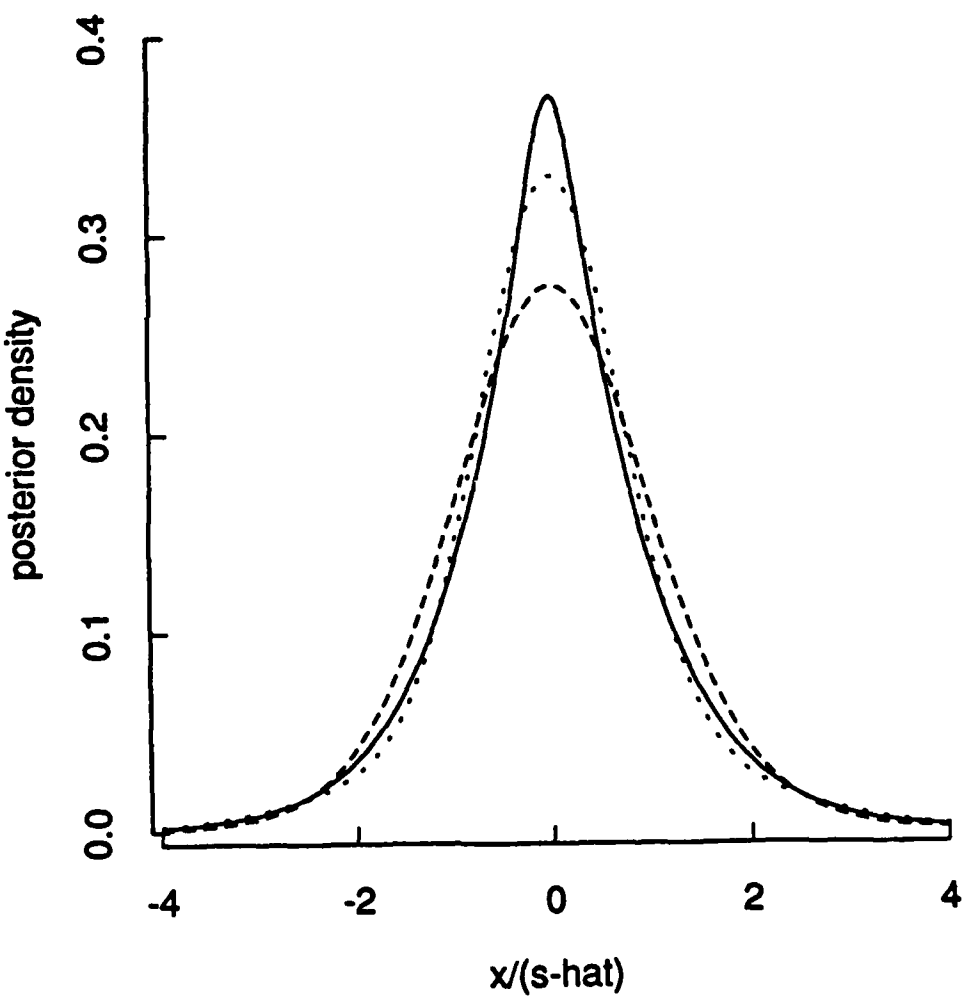


Figure 2.5b Exact and approximate posterior densities for largest expected contrast, Example 2.2, α=0.3, γ=1.25. The exact (solid curve) is obtained by direct calculation, the approximate (dashed curve) is a single t density, and the corrected approximate (dotted curve) is a t density plus a quadratic term from the Taylor's series expansion.



(c), the approximation would be exact.

The approximate density is the first term in a Taylor series expansion of the true density about the point s^2. Writing down the first three terms of this expansion, we have

$$p(\tau|y) \approx t(\tau|s^2) + \sum_{(c)} p_{(c)}(s^2_{(c)}-s^2)\, t'(\tau|s^2) + \sum_{(c)} p_{(c)}(s^2_{(c)}-s^2)^2\, t''(\tau|s^2), \tag{2.39}$$

where t' and t'' are the first and second derivatives of $t(\tau|s^2)$ with respect to s^2. The second term in the above expression is identically zero, by the definition of s^2. The third term can be rewritten to give

$$p(\tau|y) \approx t(\tau|s^2) + CV \cdot t''(\tau/s\,|\,1), \tag{2.40}$$

where the statistic

$$CV = \frac{\sum_{(c)} p_{(c)}(s^2_{(c)}-s^2)^2}{(s^2)^2} \tag{2.41}$$

and $t''(x|1)$ is t'' with scale parameter set equal to one. The quadratic term is now written as the product of the statistic CV which measures the relative variation of $s^2_{(c)}$ in the weighted sum, and $t''(x/s\,|\,1)$ which measures the nonlinearity of the t density with respect to s^2. Note that there will be a different value of CV for each of the expected contrasts τ_i.

The closeness of the t approximation may now be checked conveniently according to the size of CV, with larger values indicating a less accurate approximation. For Example 2.1 with α=0.2, γ=2.5, CV is smaller than 0.1, indicating that this is a relatively safe range as shown by the closeness of the curves in Figure 2.5a. For Example 2.2 with α=0.3 and γ=1.25, CV is approximately 0.5, and as shown in Figure 2.5b, this indicates a less accurate approximation.

A large value of CV may signal a poor approximation in two ways. First, it shows a non-negligible contribution from the quadratic term in the Taylor series expansion. Second, it may also indicate that some higher order terms are also non-negligible. Thus it is possible that including the quadratic term in the approximate density will not provide an adequate correction for a large value of CV.

To investigate which values of CV might be considered large, four samples of 50 scale parameters $s_1^2, \ldots, s_{50}^2$ were generated from gamma distributions with shape parameters chosen to give CV values of .125, .25, .375 and .5 respectively. (The randomly generated samples were then shifted to give the exact CV values given above.) Exact distributions were calculated for each sample as a weighted sum of t distributions with 15 degrees of freedom and scale parameter s_i^2, with equal weight given to each value in the sample. To examine the accuracy of probability statements about Bayesian confidence intervals, which depends mainly on the tail probabilities of the true and approximate distributions, three curves are plotted in Figure 2.6. The solid curve is the true cumulative distribution function obtained from the weighted sum, with the plot restricted to the left tail of the distribution. (The true and approximate

Figure 2.6a Weighted sum of 50 t distribution functions with CV=.125 (solid curve), approximation by single t distribution (dotted curve), and approximation by single t with quadratic correction (dashed curve).

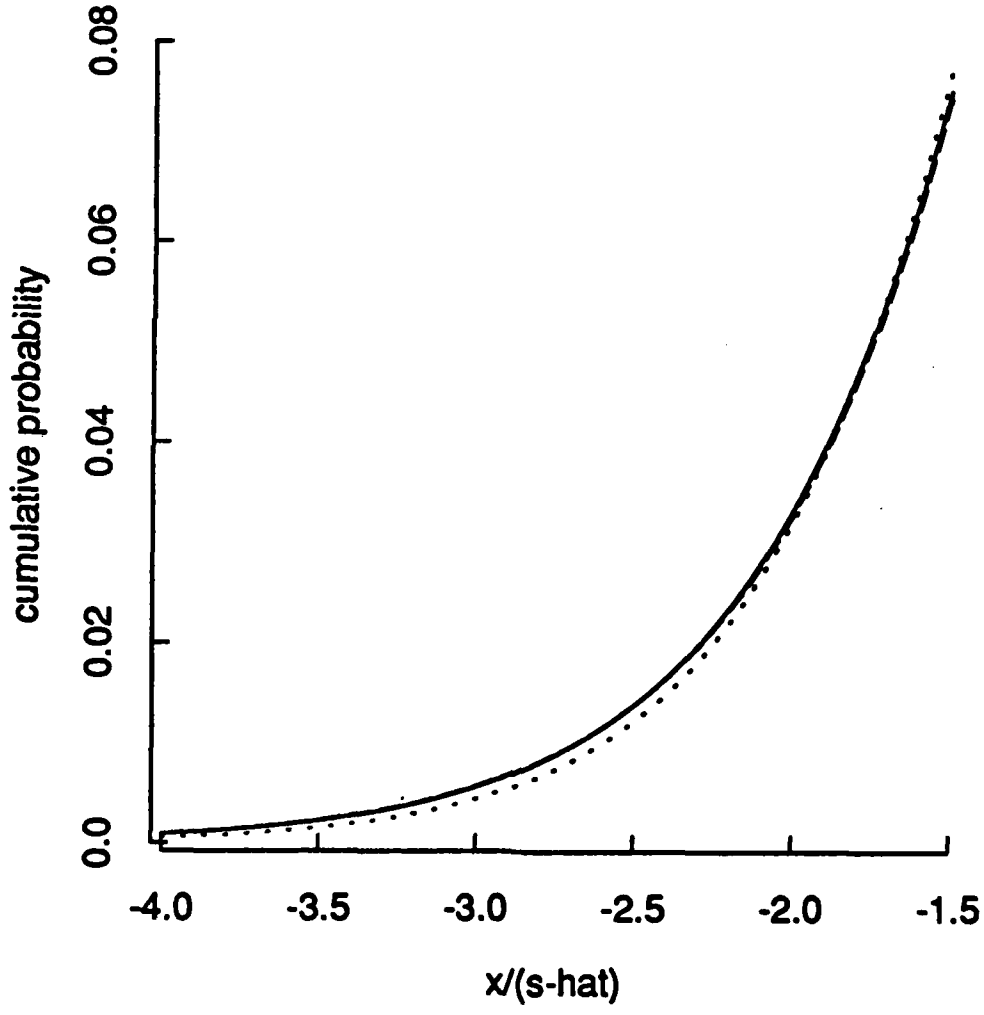


Figure 2.6b Weighted sum of 50 t distribution functions with CV=.25 (solid curve), approximation by single t distribution (dotted curve), and approximation by single t with quadratic correction (dashed curve).

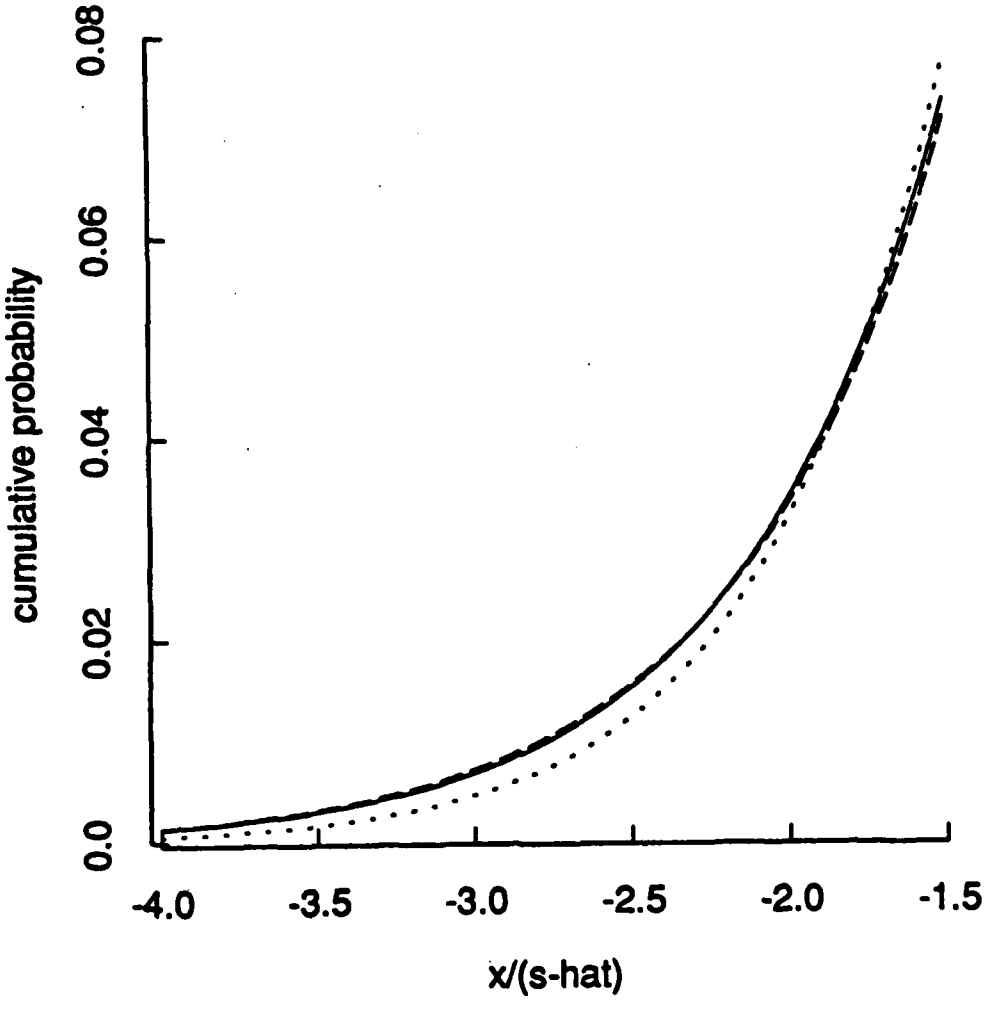


Figure 2.6c Weighted sum of 50 t distribution functions with CV=.375 (solid curve), approximation by single t distribution (dotted curve), and approximation by single t with quadratic correction (dashed curve).

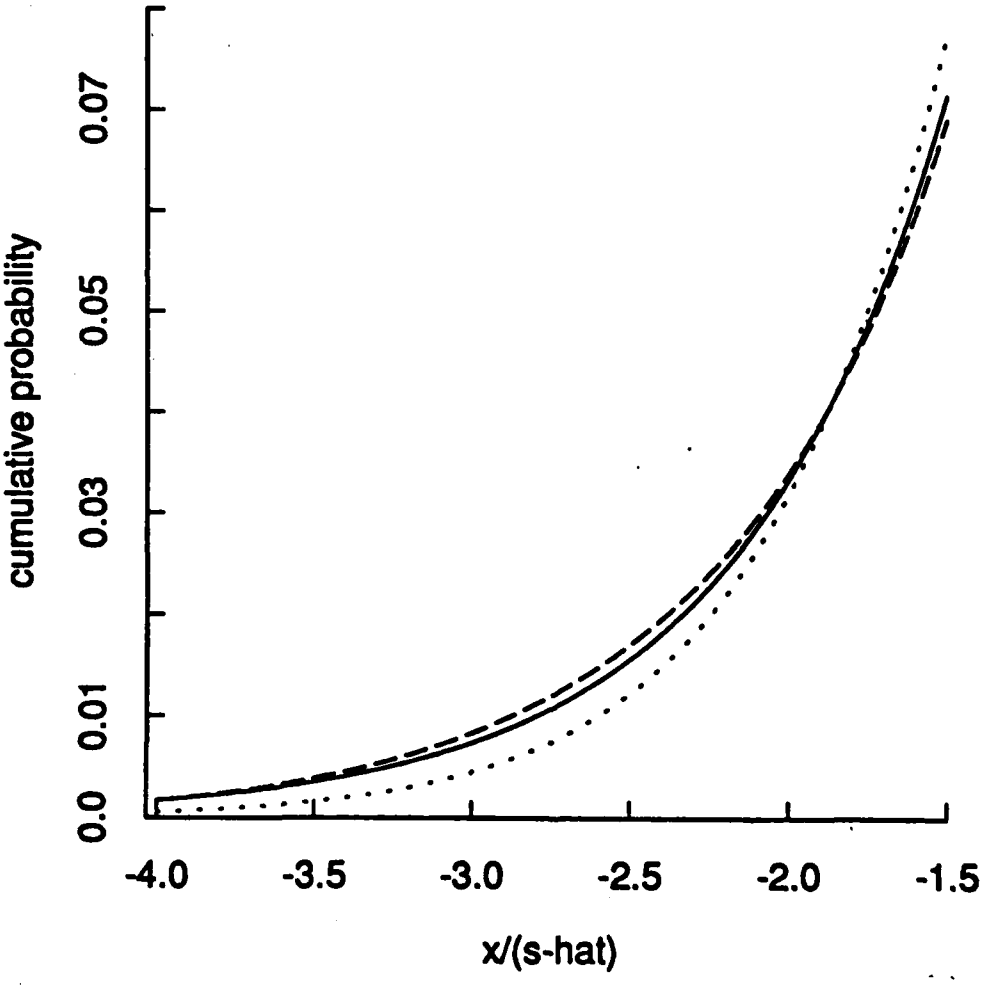
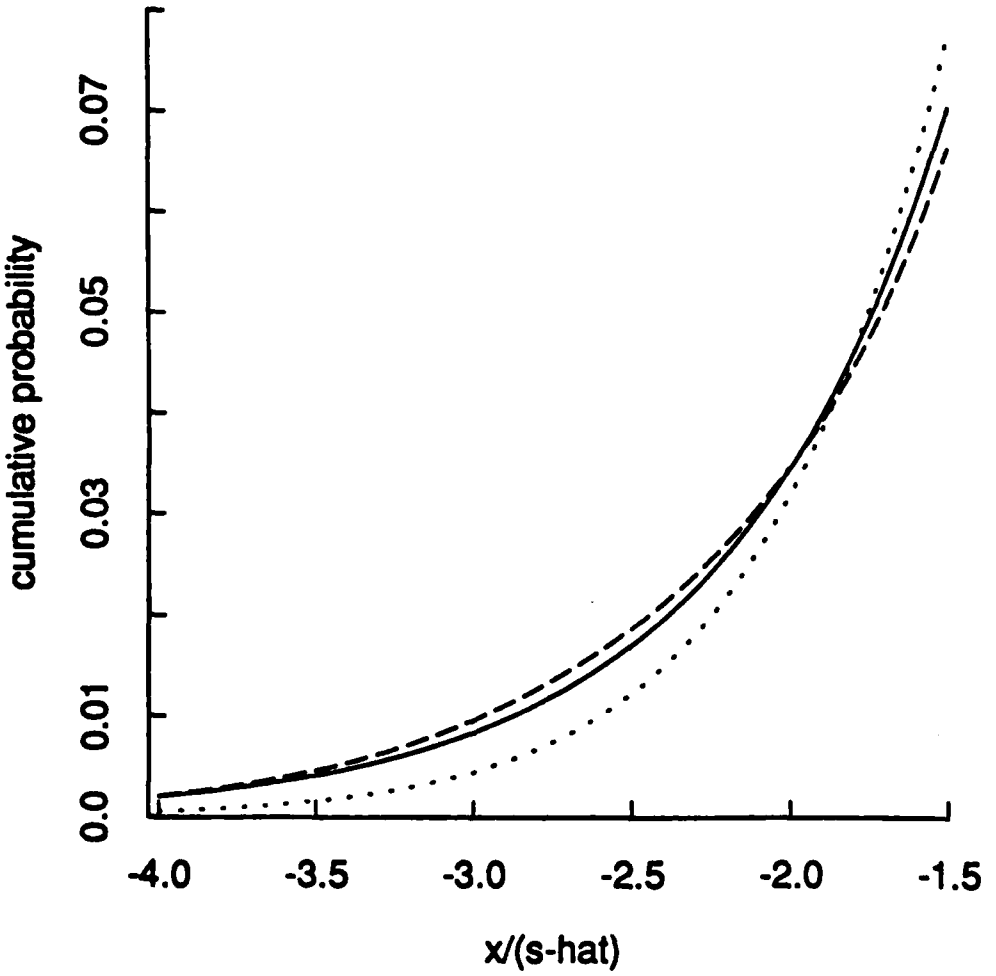


Figure 2.6d Weighted sum of 50 t distribution functions with $CV=.5$ (solid curve), approximation by single t distribution (dotted curve), and approximation by single t with quadratic correction (dashed curve).



distributions here are symmetric). The dotted curve is the approximate distribution function obtained as the c.d.f. of a single t distribution with 15 degrees of freedom and scale factor

$$s^2 = \frac{\sum_{i=1}^{50} s_i^2}{50}.$$

The dashed curve is the approximation obtained by also including the quadratic term of the Taylor series expansion.

For smaller values of CV the t approximation is fairly close to the true distribution and the quadratic correction is almost coincident with the true distribution. As CV becomes larger the t approximation becomes less and less accurate while the quadratic correction tracks the true function closely up to CV=0.5.

For extremely large values of CV the approximate density (including the quadratic term) exhibits irregular behavior. For CV larger than 1.0667 it has a local maximum in each of the tails and for CV larger than 1.6 the approximate density is negative in some regions. In general, for a weighted sum of t distributions with v degrees of freedom, the quadratic approximation develops a local maximum in each tail for CV larger than 4(v+5)/5v and becomes negative in some regions for CV larger than 4(v+3)/3v, (v=degrees of freedom).

For values of CV greater than 0.5 the quadratic approximation may give a poor estimate of the true coverage probability of a Bayesian confidence interval. However, as large values of CV are caused by different values of $s^2_{(c)}$ having significant poste-

rior probability, they correspond to situations such as Example 2.2 where the identity of active contrasts is not well-determined by the experimental data.

Table 2.4 gives quantiles of the quadratic approximation of the posterior distribution of active contrasts τ for CV between 0.0 and 0.5, degrees of freedom ν between 7 and 31 in steps of 4, and tail probabilities .005, .025 and .05. Then, for example, 95% confidence limits for an active contrast with CV=0.25 and ν=15 are $\pm 2.188(\hat{s})$.

2.4.3. Computing Details

The relevant statistics to be computed in the approximation to the posterior distribution of τ_i are $\hat{s}_i^2$ and CV_i. To utilize the factorization given in Section 2.2, expressions for these statistics must be derived in terms of $p(\sigma|\mathbf{T})$ and $p_{i|\sigma}$.

The first step is to derive the conditional posterior distribution of τ_i, given σ and given that τ_i is active,

$$p(\tau_i|\sigma,\mathbf{T},\tau_i\text{active}) = \frac{p(\sigma,\tau_i|\mathbf{T},\tau_i\text{active})}{p(\sigma|\mathbf{T},\tau_i\text{active})}. \tag{2.42}$$

Both numerator and denominator of this expression are obtained by integrating the appropriate elements of τ out of

$$p(\tau,\sigma|\mathbf{T},\tau_i\text{active}) \;\alpha \tag{2.43}$$

$$\sigma^{-n}\,\frac{1}{(2\pi(k^2-1))^{1/2}\sigma}\exp\left\{\frac{-1}{2\sigma^2}\left[(T_i-\tau_i)^2+\frac{\tau_i^2}{k^2-1}\right]\right\}\times$$

$$\prod_{j\neq i}\left[(1-\alpha)\exp\left\{\frac{T_j^2}{2\sigma^2}\right\}+\frac{\alpha}{(2\pi(k^2-1))^{1/2}\sigma}\exp\left\{\frac{-1}{2\sigma^2}\left[(T_j-\tau_j)^2+\frac{\tau_j^2}{k^2-1}\right]\right\}\right]$$

Table 2.4 Approximate quantiles of the posterior distribution of an expected contrast τ divided by its posterior standard deviation, obtained from the first three terms in a Taylor series expansion of the exact distribution. The approximate quantile is a function of the degrees of freedom and the statistic CV. Values obtained are for tail probabilities .05, .025 and .005 and degrees of freedom $n-1$ for n a multiple of four between 8 and 32.

tail probability = 0.05

	degrees of freedom						
CV	7	11	15	19	23	27	31
0.	1.895	1.796	1.753	1.729	1.714	1.703	1.696
0.05	1.891	1.793	1.750	1.726	1.711	1.700	1.692
0.10	1.887	1.789	1.747	1.723	1.707	1.697	1.689
0.15	1.884	1.786	1.743	1.719	1.704	1.693	1.686
0.20	1.880	1.782	1.739	1.715	1.700	1.690	1.682
0.25	1.875	1.778	1.735	1.711	1.696	1.685	1.678
0.30	1.871	1.773	1.731	1.707	1.692	1.681	1.673
0.35	1.866	1.769	1.726	1.702	1.687	1.676	1.668
0.40	1.861	1.764	1.721	1.697	1.682	1.671	1.663
0.45	1.856	1.759	1.716	1.692	1.676	1.666	1.658
0.50	1.850	1.753	1.710	1.686	1.670	1.659	1.652

tail probability = 0.025

	degrees of freedom						
CV	7	11	15	19	23	27	31
0.	2.365	2.201	2.131	2.093	2.069	2.052	2.040
0.05	2.373	2.211	2.141	2.103	2.079	2.062	2.050
0.10	2.382	2.221	2.152	2.114	2.090	2.073	2.061
0.15	2.391	2.231	2.163	2.126	2.102	2.085	2.073
0.20	2.400	2.242	2.175	2.138	2.115	2.098	2.086
0.25	2.410	2.254	2.188	2.152	2.128	2.112	2.101
0.30	2.421	2.267	2.202	2.166	2.143	2.128	2.116
0.35	2.432	2.280	2.217	2.182	2.159	2.144	2.133
0.40	2.443	2.295	2.232	2.198	2.177	2.162	2.151
0.45	2.455	2.310	2.249	2.216	2.195	2.181	2.171
0.50	2.467	2.326	2.267	2.235	2.216	2.202	2.192

tail probability = 0.005

	degrees of freedom						
CV	7	11	15	19	23	27	31
0.	3.500	3.106	2.947	2.861	2.807	2.771	2.744
0.05	3.545	3.157	3.000	2.916	2.863	2.827	2.801
0.10	3.591	3.208	3.054	2.971	2.920	2.884	2.859
0.15	3.637	3.259	3.108	3.027	2.976	2.942	2.917
0.20	3.683	3.310	3.161	3.082	3.032	2.999	2.974
0.25	3.728	3.360	3.214	3.136	3.087	3.054	3.030
0.30	3.772	3.410	3.265	3.188	3.139	3.107	3.083
0.35	3.817	3.458	3.314	3.238	3.190	3.157	3.134
0.40	3.860	3.504	3.362	3.286	3.238	3.205	3.182
0.45	3.903	3.549	3.408	3.331	3.283	3.251	3.227
0.50	3.944	3.593	3.451	3.375	3.326	3.293	3.269

After performing the integration we have

$$p(\tau_i \mid \sigma, \mathbf{T}, \tau_i \text{ active}) = (2\pi\phi)^{-1/2}\sigma^{-1}\exp\left\{\frac{-(\tau_i-\phi T_i)^2}{2\phi\sigma^2}\right\}, \tag{2.44}$$

a normal distribution with mean ϕT_i and variance $\phi\sigma^2$. (Recall $\phi=1-1/k^2$). In addition, the conditional posterior of σ, given τ_i is active, can be written

$$p(\sigma \mid \mathbf{T}, \tau_i \text{active}) \;\alpha\; p(\sigma \mid \mathbf{T}) p_{i\mid\sigma}\,. \tag{2.45}$$

The quantities of interest, $s_i^{\,2}$ and CV_i, are functions of the second and fourth conditional posterior moments of τ_i, given it is active. Thus they can be obtained as the corresponding second and fourth moments of the contitional posterior distribution (2.44) of τ_i given it is active and given σ, integrated against the conditional posterior distribution (2.45) of σ given τ_i is active. Doing this, the following expressions are obtained:

$$s_i^{\,2} = \frac{n-3}{n-1}\int_0^\infty \phi\sigma^2 p(\sigma \mid \mathbf{T}, \tau_i \text{active})\, d\sigma \tag{2.46}$$

$$CV_i = \frac{\dfrac{(n-3)(n-5)}{(n-1)^2}\displaystyle\int_0^\infty \phi^2\sigma^4 p(\sigma \mid \mathbf{T}, \tau_i \text{active})\, d\sigma}{(s_i^{\,2})^2} - 1 \tag{2.47}$$

2.4.4. Marginal Posterior of σ

The marginal posterior distribution of σ is obtained from the joint posterior distribution of $\{\tau,\sigma\}$ in the same way as the posterior distribution of τ was obtained. Thus $p(\sigma \mid \mathbf{y})$ is written

$$p(\sigma|y) = \sum_{(c)} p(\sigma|y,a_{(c)})p(a_{(c)}|y). \tag{2.48}$$

The posterior distribution of σ conditional on $a_{(c)}$ is obtained by integration,

$$p(\sigma|y,a_{(c)}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sigma^{-n-c-1} \exp\left\{\frac{-1}{2\sigma^2}\left[S(\tau_{(c)}) + \tau_{(c)}'\Gamma_c\tau_{(c)}\right]\right\} d\tau_{(c)}$$

$$\alpha\ \sigma^{-n}\exp\left\{\frac{\left[S(\hat{\tau}_{(c)}) + \hat{\tau}_{(c)}'\Gamma_c\hat{\tau}_{(c)}\right]}{2\sigma^2}\right\}. \tag{2.49}$$

After normalizing, the conditional posterior distribution of σ is

$$p(\sigma|y,a_{(c)}) = \left[\frac{1}{2}\Gamma\left[\frac{n-1}{2}\right]\right]^{-1}\left[\frac{(n-1)s^2_{(c)}}{2}\right]^{(n-1)/2} \times$$

$$\sigma^{-n}\exp\left\{\frac{-(n-1)s^2_{(c)}}{2\sigma^2}\right\} \tag{2.50}$$

where $s^2_{(c)}$ is defined by (2.31). The distribution of σ is a scaled inverted chi distribution with $n-1$ degrees of freedom. Analogous to the posterior distribution of τ, the complete posterior distribution of σ is a weighted sum of 2^{n-1} inverted chi distributions with different parameters $s^2_{(c)}$.

The Taylor series approach used to approximate the posterior distribution of τ was unsatisfactory for σ, possibly because of the skewness of the densities $p(\sigma|y,a_{(c)})$. However, a normal approximation to the posterior distribution of $\log(\sigma^2)$ gives a better fit and is especially convenient for obtaining Bayesian confidence limits for σ or σ^2.

Posterior moments for $\log(\sigma^2)$ are

$$E[\log(\sigma^2)|T] = \int_0^\infty \log\sigma^2 \, p(\sigma|T) d\sigma \tag{2.51}$$

$$E[(\log(\sigma^2))^2|T] = \int_0^\infty (\log\sigma^2)^2 p(\sigma|T) d\sigma \tag{2.52}$$

with the variance obtained in the usual way from these. The posterior distribution of $\log\sigma^2$ is approximated by a normal distribution with the above mean and variance. While the posterior distribution of $\log(\sigma^2)$ is not symmetric in general, the individual components $p(\log(\sigma^2)|y,a_{(c)})$ are nearly symmetric with nearly constant variance (Box and Tiao, 1973), and the log transformation also somewhat symmetrizes the distribution of the parameters $s^2_{(c)}$ over $a_{(c)}$. Thus the posterior distribution of $\log(\sigma^2)$ is much closer to being symmetric than the posteriors of σ or σ^2. Exact and approximate posterior densities for Examples 2.1 and 2.2 are given in Figure 2.7. Approximate 95% confidence limits for σ^2 are obtained by

$$\exp\left\{E[\log(\sigma^2)|T] \pm 1.96\,(\mathrm{Var}[\log\sigma^2|T])^{1/2}\right\}. \tag{2.53}$$

For Examples 2.1 and 2.2 these are

Example 2.1: (.032, .176)

Example 2.2: (.0010, .0114).

Figure 2.7a Exact posterior density (solid curve) of $\log(\sigma^2)$ and normal approximation (dotted curve), Example 2.1.

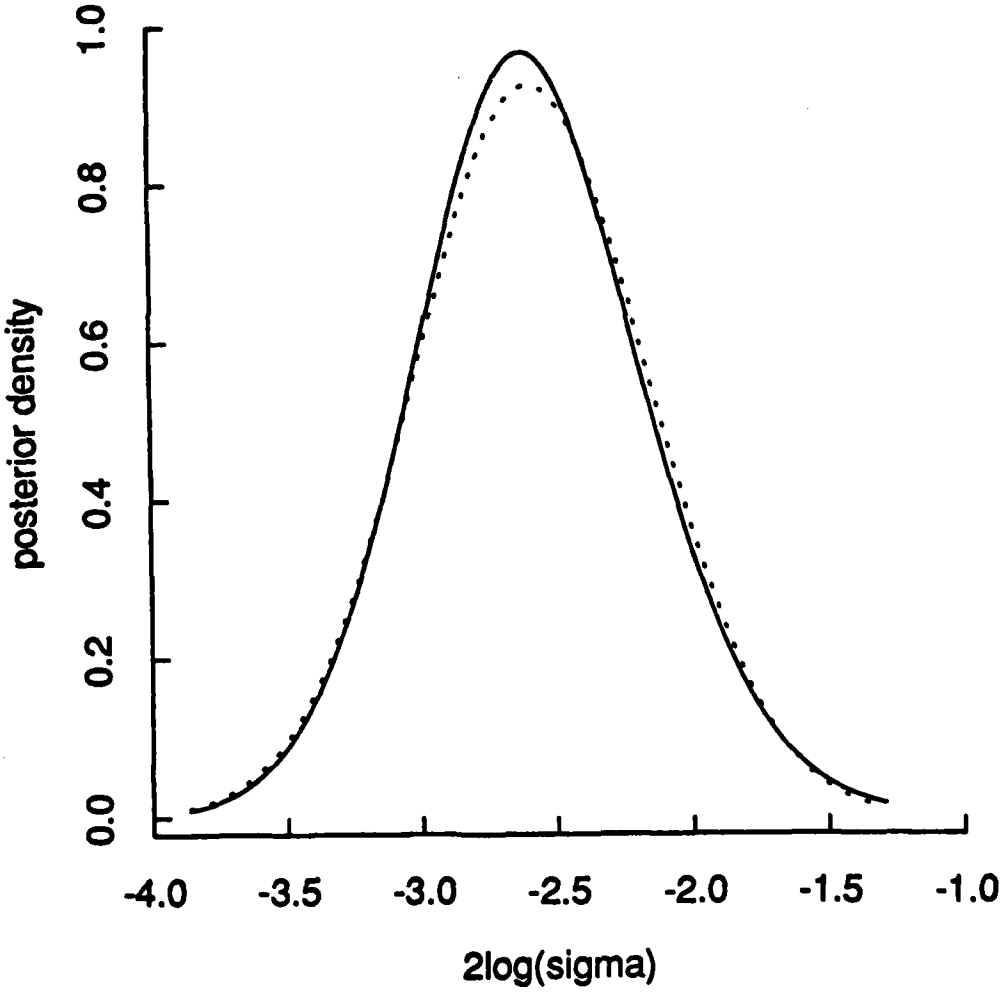
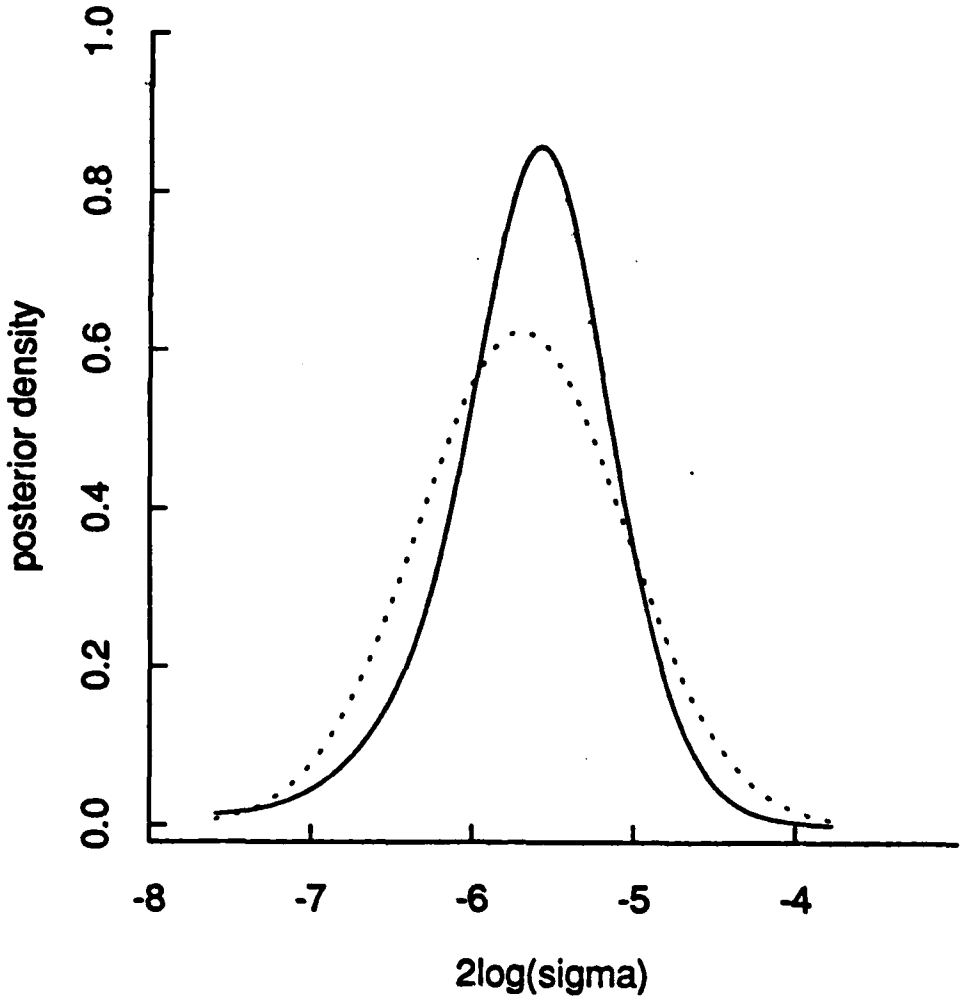


Figure 2.7b Exact posterior density (solid curve) of $\log(\sigma^2)$ and normal approximation (dotted curve), Example 2.2.



2.5. Degrees of Freedom and the Parameter k

Given the event $a_{(c)}$ that a particular combination of c contrasts is active, the posterior distribution of $\tau_{(c)}$ was shown to be a t distribution with $n-1$ degrees of freedom. And if active and inert contrasts are well determined by the experimental data, the posterior distributions of the supposed active contrasts averaged over all events $a_{(c)}$ are quite close to single t distributions with $n-1$ degrees of freedom. This is in contrast to the usual situation where the degrees of freedom are equal to n minus the number of parameters estimated. The reason for this is that the prior variance of the parameters τ was assumed to be a known constant involving k, times σ^2. Thus in the expression (2.31) for the posterior variance of $\tau_{(c)}$ given $a_{(c)}$, there are $n-1-c$ degrees of freedom for the residual sum of squares plus c degrees of freedom for the active contrasts which are scaled by the matrix Γ_c and included in the expression for the posterior variance. The extra c degrees of freedom appear because it is assumed that the squared active contrasts, scaled by a known constant, also estimate σ^2.

It was shown previously that varying k often has little effect on the posterior probabilities $\{p_i\}$. However, when there are several apparently active contrasts, and k is not well-determined in advance, changing k may have a noticeable impact on the posterior standard deviations of the contrasts. When situations such as this do occur, two possible remedies come to mind.

If the parameter k is treated as unknown, then the predictive distribution

$$p(y) = \int_{-\infty}^{\infty}\int_{0}^{\infty} p(y \mid \sigma,\tau)\, p(\tau \mid \sigma)\, p(\sigma)\, d\sigma d\tau \tag{2.54}$$

as a function of k given y can be thought of as being proportional to the posterior density of k under a locally uniform prior. Denote this function by $p(k \mid y)$, and the posterior variance of a typical τ_i given k by $s_i^2(k)$. Then an expression for the unconditional posterior variance of τ_i is given by

$$s_i^2 = \frac{\int s_i^2(k)\, p(k \mid y)\, dk}{\int p(k \mid y)\, dk}. \tag{2.55}$$

This can be approximated by choosing a few values of k, such as 5, 10 and 15, and estimating the integral by

$$s_i^2 = \frac{\sum_k s_i^2(k)\, p(k \mid y)}{\sum_k p(k \mid y)}. \tag{2.56}$$

A second method for estimating the standard error of a contrast is motivated by the usual linear model approach of specifying which coefficients are significant, and estimating variance from the residuals. The proposed variance of an observed contrast is

$$v_i^2 = \frac{\sum_{j \neq i} T_j^2 (1 - p_j)}{n - 1 - \sum_j p_j} \tag{2.57}$$

which is analogous to the linear model approach, but accounts for the indeterminacy of which contrasts are active. When all probabilities are close to either zero or one,

the estimate agrees closely with the standard approach. When there are probabilities closer to 1/2, the estimate is a weighted average of the standard estimates obtained by either including or excluding the in-between contrasts. This estimator of variance is insensitive to changes in k as long as the probabilities are also insensitive.

These two methods are compared with the posterior variance derived in the previous section 2.4.2, using the data of Example 2.1, in particular the apparently active contrast from column 3 of the design. The posterior standard deviation of this contrast was computed with α fixed at 0.2 and k taking the values 5, 10 and 15. For each combination of α and k the estimate v_3 was also computed. The unconditional posterior standard deviation s_3 was computed over the three values of k as indicated by the formula (2.56). The posterior standard deviations $s_3(k)$ of column contrast 3 for k=5, 10 and 15 were 0.640, 0.534 and 0.517, respectively. The estimates v_3 were 0.583, 0.571 and 0.573, respectively. The marginal weighted estimate s_3 was 0.570. The estimates v_3 are much less sensitive to changes in k than the regular posterior moments, and are also close to the weighted average s_3.

The marginal weighted average estimate s_i given by (2.55) is the "correct" estimate from the Bayesian viewpoint when precise knowledge of k is not available. However, because the posterior density of k reaches a maximum at a value close to the ratio of the mean squared active contrast over the mean squared inert contrast, and the posterior standard deviation of τ_i at this value of k is approximately equal to the mean squared inert contrast, the integral (2.55) will also be close to the mean squared inert contrast. The estimator v_i is basically also equal to the mean squared inert con-

trast and thus estimates the unconditional standard deviation s_i^* without repeating the calculations for several values of k, making it a convenient and less sensitive estimate of the standard error of an active contrast.

Although the integration with respect to k is not done analytically, it is probably safe to say the form of the posterior distribution of τ_i has changed from the discrete mixture of t distributions with $n-1$ degrees of freedom, given k. Because of the analogy with the usual linear model, it was thought reduction of the degrees of freedom to $n-1-\sum p_j$ would help correct for the new form of the distribution. However, comparing nominal and actual coverage probabilities of confidence intervals for the apparently active contrasts of Example 2.1, the closest agreement was attained when approximating the posterior distribution by a t density with $n-1$ degrees of freedom.

2.6. Replication and Blocking

2.6.1. Replication

The model described here may be applied to any factorial experiment, unreplicated or otherwise. Although it provides a method of analysis for the case when there are no true replicate runs, as a general method for analyzing fractional factorials the model can be applied successfully when there are replicate runs. For example, consider the special case of an $n \times n$ design $\mathbf{X}$ with the properties described in Section 1.2, with m independent observations at each point of the design. In general it is assumed that the experiment is run in m blocks of n runs, with each block consisting of one replication of the design $\mathbf{X}$, each block having a different mean, with no interaction between blocks and factors. The analysis can be reduced to the case where each block mean is assumed the same, to be applied when the replicates are not run in blocks. The situation where blocking is done within the design $\mathbf{X}$ is described in Section 2.6.3.

2.6.2. Joint Posterior for the Replicated Design

Let $\mathbf{y}_j$ be the vector of observations for the jth block. Again make the one-to-one transformation

$$\mathbf{T}_j = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_j, \quad j = 1,\ldots, m.$$

The sampling distribution of $\mathbf{T} = (\mathbf{T}_1, \ldots, \mathbf{T}_m)$ is

$$p(\mathbf{T}\,|\,\tau,\sigma)\ \alpha\ \sigma^{-mn}\prod_{i=0}^{n-1}\exp\left\{\frac{-\sum_{j=1}^{m}(T_{ij}-\tau_i)^2}{2\sigma^2}\right\}. \tag{2.58}$$

Utilizing the same prior distributions as given in Section 2.1, the joint posterior distribution of $\{\tau,\sigma\}$ is

$$p(\tau,\sigma\,|\,\mathbf{T})\ \alpha\ \sigma^{-mn-1}\exp\left\{\frac{-\sum_{j=1}^{n-1}(T_{0j}-\tau_{0j})^2}{2\sigma^2}\right\}\times \tag{2.59}$$

$$\prod_{i=0}^{n-1}\left[\exp\left\{\frac{-\sum_{j=1}^{m}(T_{ij}-\tau_i)^2}{2\sigma^2}\right\}\frac{\alpha}{((2\pi)(k^2-1))^{1/2}\sigma}\exp\left\{\frac{-\tau_i{}^2}{2(k^2-1)\sigma^2}\right\}\right.$$

$$\left.+\,(1-\alpha)\exp\left\{\frac{-1}{2\sigma^2}\sum_{j=1}^{n}T_{ij}{}^2\right\}\right].$$

The quantities

$$\sum_{j=1}^{m}(T_{ij}-\tau_i)^2,\quad \sum_{j=1}^{m}T_{ij}{}^2$$

can be decomposed to give

$$\sum_{j=1}^{m}(T_{ij}-\tau_i)^2=m(\bar{T}_i-\tau_i)^2+S_i \tag{2.60}$$

and

$$\sum_{j=1}^{m}T_{ij}{}^2=m(\bar{T}_i)^2+S_i \tag{2.61}$$

with

$$S_i = \sum_{j=1}^{m} (T_{ij} - \bar{T}_i)^2 . \tag{2.62}$$

Thus the posterior distribution of $\{ \tau, \sigma \}$ can be written

$$p(\tau,\sigma|T) \;\alpha\; \sigma^{-mn-1} \exp\left\{ \frac{-\sum_{j=1}^{m} (T_{0j} - \tau_{0j})^2}{2\sigma^2} \right\} \exp\left\{ \frac{-\sum_{i=1}^{n-1} S_i}{2\sigma^2} \right\} \times \tag{2.63}$$

$$\prod_{i=1}^{n-1} \left[\exp\left\{ \frac{-m(\bar{T}_i - \tau_i)^2}{2\sigma^2} \right\} \frac{\alpha}{((2\pi)(k^2-1))^{1/2}\sigma} \exp\left\{ \frac{-\tau_i^2}{2(k^2-1)\sigma^2} \right\} \right.$$

$$\left. + (1-\alpha) \exp\left\{ \frac{-m\bar{T}_i^2}{2\sigma^2} \right\} \right] .$$

This is of the same form as that derived for the unreplicated case, with the variance reduced to σ^2/m by replication, the non-informative prior distribution for σ replaced by an informative inverted χ distribution, and k^2 defined to be $mn\gamma^2+1$. For the rest of this section σ^2 will refer to the reduced variance, which is $1/mn$ times the original error variance, or $1/m$ times the variance of an observed contrast from an unreplicated n-run experiment. The two cases considered here are:

1. For unequal, unknown block effects, the prior estimate of σ^2 is

$$s^2 = \frac{\sum_{i=1}^{n-1} S_i}{m(m-1)(n-1)} \tag{2.64}$$

with $(m-1)(n-1)$ degrees of freedom.

2. For equal block effects (equivalently no block effects, but one overall mean), the prior estimate of σ^2 is

$$s^2 = \frac{\sum_{i=0}^{n-1} S_i}{m(m-1)n} \tag{2.65}$$

with $(m-1)n$ degrees of freedom.

In each of the above cases the extra divisor of m in the formula for s^2 appears because of the previously mentioned m-fold variance reduction.

Thus the Bayesian model applied to a replicated experiment is equivalent to the unreplicated analysis on the observed contrasts averaged over the replicates, with a prior estimate of variance obtained as the variance of the observed contrasts between replicates. It is important to note that this result depends upon the orthogonality among contrasts and replicates obtained by repeating each point the same number of times. Unequal replication does not lead to this simplification.

2.6.3. Blocking

In situations where it is not possible to complete all n runs of the design in the same day, or with the same batch of raw material, or with the same technician, etc., it is common to run the experiment in blocks, associating effects of supposedly lesser importance (high order interactions) with block differences (see, e.g., Cochran and Cox, 1957, p. 183; Box, Hunter and Hunter, 1978, p. 336). I describe below how to deal with block effects in the proposed analysis.

For a first attempt one might pretend that those contrasts associated with block effects still come from the normal population implied by the prior distribution (1.3), and are active with probability α. This is equivalent to ignoring the fact that some of the contrasts are biased by block differences, and continuing with the analysis as usual. The resulting posterior distribution of σ may, however, be contaminated by these possibly inflated contrasts. A safer method is to assume there will always be some effect due to blocking, and assign a noninformative prior to the magnitudes of these effects. The result is that contrasts associated with block differences do not enter into the calculation of posterior probabilities and related statistics for the contrasts of interest, and they are ignored just as the grand mean was ignored in the previous analysis.

Suppose then, that the design $\mathbf{X}$ is replicated m times, with b blocks within each replicate of $\mathbf{X}$, and the same columns are associated with blocks in each replicate. Assuming no interaction between blocks and factors, there are $n-b$ contrasts for which posterior probabilities will be computed. If $m>1$, there will be $(m-1)(n-b)$ additional degrees of freedom for estimating variance. Following the same steps as in the previous section, the posterior distribution of σ and the contrasts of interest τ is

$$p(\tau,\sigma|\mathrm{T}) = \sigma^{-m(n-b)-1} \exp\left\{\frac{-m\sum_{i=1}^{n-b} S_i}{2\sigma^2}\right\} \times \tag{2.66}$$

$$\prod_{i=1}^{n-b}\left[\exp\left\{\frac{-m(\bar{T}_i-\tau_i)^2}{2\sigma^2}\right\}\frac{\alpha}{(2\pi(k^2-1))^{1/2}\sigma}\exp\left\{\frac{-\tau_i{}^2}{2(k^2-1)\sigma^2}\right\}\right.$$

$$\left. + (1-\alpha)\exp\left\{\frac{-m\bar{T}_i{}^2}{2\sigma^2}\right\}\right].$$

where

$$S_i = \frac{\sum_{j=1}^{m}(T_{ij}-\bar{T}_i)^2}{m}.$$

Again, this is of the same form (2.12) as for the unreplicated design, with the noninformative prior for σ replaced by the informative inverted χ distribution.

2.6.4. Dependence of α and k on m,n

The parameters k, γ are related by the equation

$$k^2 = mn\gamma^2 + 1 \tag{2.67}$$

and it is clear that as m or n increases, either k must increase, or γ must decrease, or both. It has been an accepted notion in statistical analysis to compare an estimator with its own variance to determine if it is "significant" or not. In this case the variance of an observed contrast is σ^2/mn, so that as m and n increase, smaller and smaller contrasts should be declared active in comparison with this variance. This implies

that the prior variance of an active contrast should decrease as a function of m and n. Specifically, if the prior belief is that active contrasts should be a certain size relative to σ^2/mn, then the prior variance should be a constant times σ^2/mn. Thus γ^2 is proportional to a constant over mn. This implies that the parameter k should be independent of m and n, by equation (2.67). Thus the range for k given in Section 2.3.1 will serve as a reasonable guide for general m, n.

Regarding α, as the number of replicates m increases for fixed n, the increasing number of expected active contrasts should be reflected in a larger value of α. For example, if it is believed that the inherent noise in a process will make it impossible to uncover any active contrasts in an unreplicated experiment, replication will serve to decrease the noise in the observed contrasts and increase the frequency of detectable large values.

2.6.5. An Example

To illustrate, the following example is taken from Barnett and Mead (1956).

Example 2.3 The authors wished to study the effect of variations in four operating factors, pH (P), aluminum reagent (A), carbon slurry (C) and barium chloride (B), on the efficiency of a decontamination process for removal of radioactivity from liquid wastes. They chose to run a 2^4 full factorial design, twice replicated. Because only eight of the sixteen factor combinations could be completed in one day, each replicate was run in two blocks of eight runs, and the four-factor interaction **PACB** was confounded with block differences within each replicate. The design, observations and calculated contrasts are given in Table 2.5.

Table 2.5 Design matrix, observations, and observed contrasts for Example 2.3, a twice-replicated 2^4 full factorial experiment run in two blocks of 16 runs, from Barnett and Mead (1956). The CABP contrast is confounded with block differences.

	factors				response	
run	C	A	B	P	y_1	y_2
1	-	-	-	-	881	834
2	+	-	-	-	650	494
3	-	+	-	-	191	257
4	+	+	-	-	183	193
5	-	-	+	-	289	178
6	+	-	+	-	188	163
7	-	+	+	-	225	370
8	+	+	+	-	135	156
9	-	-	-	+	1180	1193
10	+	-	-	+	1039	1146
11	-	+	-	+	466	890
12	+	+	-	+	781	775
13	-	-	+	+	298	273
14	+	-	+	+	238	254
15	-	+	+	+	420	429
16	+	+	+	+	350	389

column (effect)	Block 1 contrast	Block 2 contrast	column (effect)	Block 1 contrast	Block 2 contrast
0(mean)	469.6	499.6	8(P)	126.9	169.0
1(C)	-24.1	-53.4	9(CP)	29.6	25.8
2(A)	-125.8	-67.3	10(AP)	33.5	19.4
3(CA)	42.5	-0.8	11(CAP)	13.3	-10.4
4(B)	-201.8	-223.1	12(BP)	-68.3	-109.3
5(CB)	-16.0	17.4	13(CBP)	-22.0	-4.5
6(AB)	140.4	126.8	14(ABP)	10.4	-6.1
7(CAB)	-42.4	-26.8	15(CABP)	-15.9	32.6

Barnett and Mead used analysis of variance to analyze the results of the experiment. Contrasts of magnitude larger than 94 were found to be significant at the nominal .01 level, and those larger than 68 at the .05 level (with no correction for selection). Thus main effects **P**, **B** and **A** and the **PB** and **BA** interactions were judged significant at the .01 level, and the main effect **C** and the **BAC** interaction were judged significant at the .05 level.

The Bayesian posterior probabilities of the contrasts were computed with α=0.2 and k=10 and are presented in Table 2.6a, with a prior estimate of the standard deviation of an observed contrast being 30.42 with $(m-1)(n-b)$=14 degrees of freedom. For comparison the posterior probabilities and related statistics were computed pretending the values of the observed contrasts were obtained from an unreplicated experiment. These values are presented in Table 2.6b.

The five largest contrasts, declared significant at the .01 level by Barnett and Mead, all have posterior probabilities close to one. The two intermediate contrasts have posterior probabilities of .261 and .174. One could reasonably conclude for this example that the five largest contrasts are almost certainly measuring real effects, and the next two largest are also possibly active and should be considered. Of course, in practice, the conclusions will depend on the objective of the experiment.

Comparing the results to those obtained by pretending the contrasts were calculated from an unreplicated experiment, the posterior probabilities are in fairly close agreement. However, the values for the diagnostic statistics $\partial p/\partial\alpha$, $\partial p/\partial k$ and CV are all much greater for the "unreplicated" analysis. This agrees with intuition. The

Table 2.6a Posterior probabilities, standard errors (conditional on being active), *CV* values, and derivatives for Example 2.3, with α=0.2 and k=10 and a prior estimate of σ of 30.42 with 14 degrees of freedom obtained from replicates.

effect	contrast	post.prob.	se|active	CV	dp/α	50dp/dk
none		0.050				
C	-77.5	0.261	31.7	0.01	1.37	-0.29
A	-193.0	0.998	34.3	0.02	0.02	0.00
AC	41.8	0.051	33.3	0.02	0.32	-0.19
B	-424.9	1.000	34.4	0.02	0.00	0.00
BC	1.4	0.024	34.4	0.02	0.15	-0.12
BA	267.1	1.000	34.4	0.02	0.00	0.00
BAC	-69.1	0.174	31.9	0.02	1.03	-0.30
P	295.9	1.000	34.4	0.02	0.00	-0.00
PC	55.4	0.089	32.5	0.02	0.57	-0.26
PA	52.9	0.079	32.7	0.02	0.51	-0.25
PAC	2.9	0.024	34.4	0.02	0.15	-0.12
PB	-177.5	0.995	34.3	0.01	0.04	0.01
PBC	-26.5	0.033	33.9	0.02	0.20	-0.15
PBA	4.3	0.025	34.4	0.02	0.15	-0.12

Table 2.6b Posterior probabilities, standard errors (conditional on being active), *CV* values, and derivatives for Example 2.3, with α=0.2 and k=10, pretending the design was not replicated.

effect	contrast	post.prob.	se|active	CV	dp/dα	50dp/dk
none		0.050				
C	-77.5	0.137	45.9	3.99	1.62	-0.18
A	-193.0	0.711	44.5	1.47	4.59	-1.19
AC	41.8	0.039	66.7	2.58	0.35	-0.16
B	-424.8	0.916	63.7	1.48	2.06	0.05
BC	1.4	0.024	78.4	1.86	0.15	-0.12
BA	267.1	0.796	50.2	1.41	4.01	-1.32
BAC	-69.1	0.098	49.8	3.86	1.16	-0.18
P	295.9	0.822	52.7	1.42	3.69	-1.28
PC	55.4	0.058	58.2	3.23	0.62	-0.17
PA	52.9	0.053	59.8	3.10	0.56	0.02
PAC	2.9	0.024	78.3	1.86	0.15	-0.12
PB	-177.5	0.683	43.4	1.53	4.64	0.12
PBC	-26.5	0.029	73.8	2.12	0.21	-0.14
PBA	4.3	0.024	78.3	1.87	0.15	-0.12

derivatives with respect to α and k measure dependence on the choice of these prior parameters, and dependence on the prior would be expected to decrease as additional data were collected. The statistic CV can be interpreted as measuring the sharpness of the posterior distribution of σ, which is also enhanced by adding replicate observations.

2.7. Conclusions

The incorporation of assumptions into the normal theory model used for analyzing factorial and fractional factorial experiments has led to a more formal analog to the normal plot of Daniel (1959). Assessing the column contrasts according to their corresponding posterior probabilities, which can be presented graphically, is intuitively appealing. The method does not suffer the computational limitations usually associated with such elaborated models, due to the alternative Bayes factorization presented in Section 2.2. Standard errors for supposed active contrasts are easily obtained. The versatility of the analysis is also appealing: it can be used when designs are replicated and blocked, or when certain columns are assumed to be inert a priori. Overall, it provides an interesting and exciting new analysis for factorial experiments.

CHAPTER 3

IDENTIFICATION OF ACTIVE FACTORS

3.1. Introduction

As discussed in Chapter 1, the historical approach to the analysis of unreplicated factorial experiments has been to identify column contrasts which are too large to attribute to noise (see, e.g., Daniel, 1959; Box, Hunter and Hunter, 1978, p. 329). The Bayesian analysis proposed in the previous chapter addresses this problem. Once active contrasts have tentatively been identified, it remains to determine which combination(s) of the experimental variables are most likely responsible for the large observed contrasts. In this chapter I propose a formal Bayesian analysis, analogous to the one described in Chapter 2, for the problem of identifying the active factors as opposed to active contrasts.

Two basic guidelines for interpretation of fractional factorials given by Box and Hunter (1961), and restated in Section 1.3, are a) significant interactions are more likely to occur between variables which have large main effects, and b) main effects are usually larger than two-factor interactions, which are larger than three-factor interactions, etc. These guidelines are formalized in the model proposed in this chapter. The first of these guidelines is modeled by considering only column combinations corresponding to experimental factors and interactions among those factors. To follow the second guideline, separate values of the parameter k for main effects

and two-factor interactions can be specified, with k_1 for main effects larger than k_2 for two-factor interactions. Three-factor interactions can either be assumed to have the same (or smaller) variance as two-factor interactions, or they can be assumed to be inert.

3.2. The Model

The following modifications to the Bayesian model introduced previously are needed. It is assumed that factors will be active in producing main effects and interactions with prior probability α. In general this value of α will be different from the value used in the previous chapter to describe the frequency of active columns. Let $a_{(f)}$ be the event that a particular combination of f factors is active. Let $\mathbf{X}_{(f)}$ be the matrix of columns of $\mathbf{X}$ corresponding to the active effects of $a_{(f)}$ (including interactions). For example, if $a_{(f)}$ is the event that factors 1 and 2 are active, $\mathbf{X}_{(f)}$ would contain columns for the main effects of, as well as a column for the two-factor interaction between, factors 1 and 2. Likewise let $\tau_{(f)}$ be the vector of true effects under $a_{(f)}$. The sampling distribution of the vector of observations $\mathbf{y}$, given $a_{(f)}$, is

$$p(\mathbf{y} \mid a_{(f)},\sigma,\tau_{(f)}) \;\alpha\; \sigma^{-n} \exp\left\{\frac{-1}{2\sigma^2}(\mathbf{y}-\mathbf{X}_{(f)}\tau_{(f)})'(\mathbf{y}-\mathbf{X}_{(f)}\tau_{(f)})\right\}. \quad (3.1)$$

The elements of $\tau_{(f)}$ are assumed to have independent, but not necessarily identical, prior normal distributions as before. In particular, it will be assumed that elements of $\tau_{(f)}$ which are main effects will have prior distributions with mean 0 and variance $\gamma_1^2\sigma^2$, and those elements which are two-factor interactions will have mean 0 and variance $\gamma_2^2\sigma^2$. And, though this assumption is not necessary, for ease of illustration

it will be assumed that interactions between three or more factors are inert. A noninformative prior distribution is assumed again for the overall mean τ_0 and $\log(\sigma)$, so that $p(\tau_0,\sigma) \propto 1/\sigma$ where the likelihood is appreciable. The posterior probability of the event $a_{(f)}$ can then be written

$$p_{(f)} = p(a_{(f)}|y) = C\left[\frac{\alpha}{1-\alpha}\right]^f \gamma_1^{-f}\gamma_2^{-f(f-1)/2} \times \frac{|X_0'X_0|^{1/2}}{|\Gamma_f + X_{(f)}'X_{(f)}|^{1/2}} \left[\frac{S(\hat{\tau}_{(f)}) + \hat{\tau}_{(f)}'\Gamma_f\hat{\tau}_{(f)}}{S(\hat{\tau}_{(0)})}\right]^{-(n-1)/2}, \tag{3.2}$$

where

$$\hat{\tau}_{(f)} = \left[\Gamma_f + X_{(f)}'X_{(f)}\right]^{-1} X_{(f)}'y ,$$

Γ_f is the diagonal matrix with the appropriate diagonal elements (the (i,i) element is $1/\gamma_1^2$ if the ith element of $\tau_{(f)}$ is a main effect, $1/\gamma_2^2$ if an interaction), and $S(\hat{\tau}_{(f)})$ is the residual sum of squares obtained when estimating $\tau_{(f)}$ by $\hat{\tau}_{(f)}$. (Allowance for possible higher-order interactions can be made by appropriate redefinition of $X_{(f)}$ and $\tau_{(f)}$ and the exponent of γ_2, or introduction of a third parameter γ_3). Making the transformation

$$k_j^2 = n\gamma_j^2 + 1 ,$$

the probability $p_{(f)}$ can be rewritten

$$p_{(f)} \propto \left[\frac{\alpha}{1-\alpha}\right]^f k_1^{-f} k_2^{-f(f-1)/2} \times \left[1 - \phi_1 \frac{\mathbf{T}_{m(f)}'\mathbf{T}_{m(f)}}{\mathbf{T}'\mathbf{T}} - \phi_2 \frac{\mathbf{T}_{i(f)}'\mathbf{T}_{i(f)}}{\mathbf{T}'\mathbf{T}}\right]^{-(n-1)/2}, \tag{3.3}$$

where $\mathbf{T}_{m(f)}$ is the vector of observed contrasts which are main effects under $a_{(f)}$, $\mathbf{T}_{i(f)}$ is the vector of contrasts which are interactions under $a_{(f)}$, and $\phi_j = 1-1/k_j^2$.

The probabilities $p_{(f)}$ can be accumulated to compute the marginal posterior probability p_j^* that factor j is active,

$$p_j^* = \sum_{(f):\, j \text{ active}} p_{(f)}. \tag{3.4}$$

It was shown in Chapter 2 that to compute the posterior probabilities $\{p_i\}$ that particular columns are active it was not necessary to sum probabilities over all possible combinations of active columns, but rather these could be computed via numerical integration at a considerable savings in computing time. The same is not true of the probabilities $\{p_i^*\}$, which must be computed by direct enumeration over all events $a_{(f)}$. However, for moderate experiments with fewer than 15 factors, the computations are quite manageable.

For application to fractional factorials the above definitions are consistent so long as f is restricted to be smaller than the design resolution. In the next section I give a a natural extension of the model which can be used when this assumption is too restrictive.

3.2.1. Relaxing the Bound on f

The assumption that the number of active factors is less than the design resolution will be unreasonable for fractional factorial designs of low resolution. In Example 2.1, eight factors were screened using a 2^{8-4} design of resolution four. In this situation it was unknown which of eight factors was important, and thus unlikely that the experimenter could be sure that at most three were active. A natural extension of the above ideas to allow relaxation of the bound on f is given for the 2^{k-p} designs and easily extended to more general designs.

Consider a combination of f factors denoted by $a_{(f)}$, where f is greater than or equal to the design resolution R. Suppose there is confounding among the possible main effects and interactions of $a_{(f)}$, i.e., there are column contrasts which estimate more than one of the possible effects under $a_{(f)}$. (It is possible to have combinations of f factors with $f \geq R$ for which there is no confounding, and no modification is necessary for these). For those columns which estimate more than one effect define the corresponding element of $\tau_{(f)}$ to be the linear combination of effects estimated by that column. The prior distribution of such elements of $\tau_{(f)}$ will still be independent and normal, but with variance equal to the sum of the variances for the individual components. For example, if a particular column contrast T_i estimated the sum of two two-factor interactions, the prior variance of τ_i would be $2\gamma_2{}^2\sigma^2$. All further computations proceed as usual given this modification of the prior distribution of $\tau_{(f)}$. For example, consider a combination of four factors which are confounded from the 2^{8-4} design of Example 2.1. (The Hadamard product of the columns of any three of the

factors gives the column of the remaining factor). There will be three column contrasts each of which estimates the sum of two two-factor interactions among the four factors. The posterior probability of this combination is given by (assuming interactions between three or more factors to be inert)

$$p_{(f)} \propto \left[\frac{\alpha}{1-\alpha}\right]^4 k_1^{-4}(2^{1/2}k_2)^{-3} \times \left[1-\phi_1\frac{\mathbf{T}_{m(f)}'\mathbf{T}_{m(f)}}{\mathbf{T}'\mathbf{T}}-\left\{1-\frac{1}{2k_2^2}\right\}\frac{\mathbf{T}_{i(f)}'\mathbf{T}_{i(f)}}{\mathbf{T}'\mathbf{T}}\right]^{-(n-1)/2}, \tag{3.5}$$

where $\mathbf{T}_{i(f)}$ is the vector of contrasts each estimating a sum of two interactions.

3.3. Prior Parameters

To estimate plausible values for α, k_1 and k_2, the published examples given in Table 2.1 are reexamined. For each example, α is estimated by the proportion of factors declared active by the authors, k_1^2 is estimated by the mean squared main effect among active factors over the mean squared inert contrast, and k_2^2 is estimated by the mean squared two-factor interaction among active factors over the mean squared inert contrast. In this context not all active contrasts will be large, although all inert contrasts should be small. The estimated values of α, k_1 and k_2 are presented in Table 3.1.

For those examples which are full factorials and the one which is a half-fraction, at least half of the variables were declared active. For the more highly fractionated designs, of course, a much smaller proportion of the variables were found to be active.

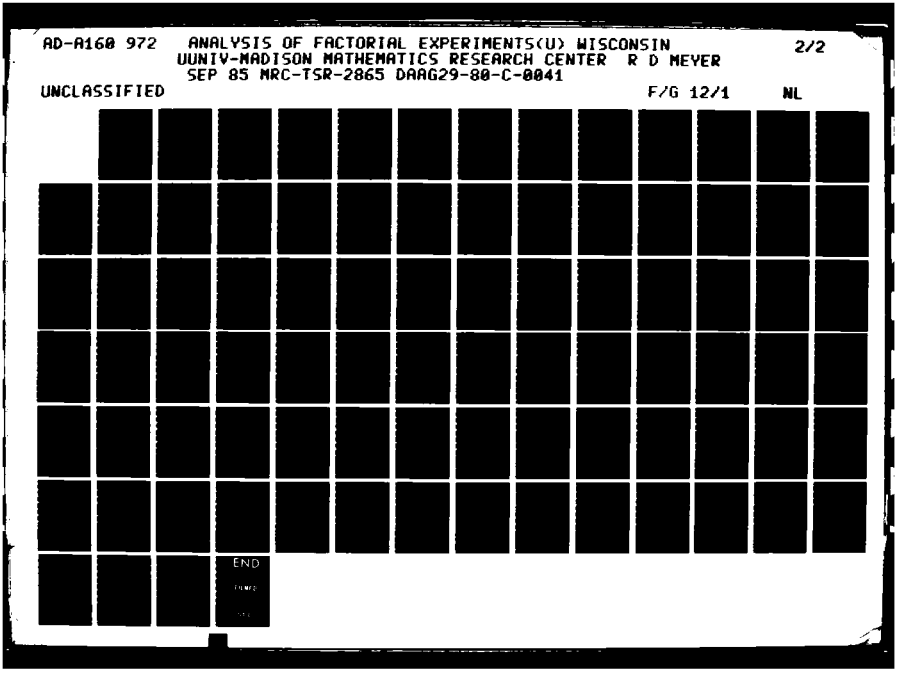
Table 3.1 Estimated values of α, k_1 for main effects and k_2 for interactions for the modified Bayesian model, from published examples of two-level experiments taken from Box, Hunter and Hunter (1978), Davies ed. (1954), Daniel (1976), Bennett and Franklin (1954), Johnson and Leone (1964), and Taguchi and Wu (1980). In Daniel's example the analysis is conducted after making a log transformation in the response.

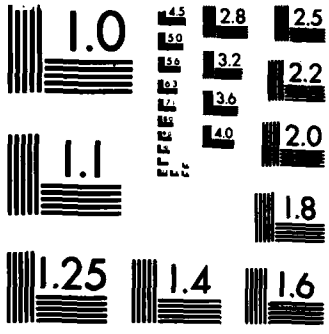
Example	n	fraction	α	k_1	k_2
BHH p. 398	16	1/16	.38	9.3	6.5
BHH p. 327	16	1	.75	15.2	2.7
BHH p. 378	32	1	.60	11.8	8.9
Davies p. 274	16	1	.50	1.9	2.5
Davies p. 462	16	1/2	.80	7.6	2.2
Daniel p. 71	16	1	.75	13.0	1.0
BF p. 557	16	1	.75	26.8	6.3
JL p. 183	32	1	.60	3.0	1.1
JL p. 196	16	1	.75	11.9	1.5
TW p. 69	16	1/32	.22	9.5	<<1.0
Average		low	.69	11.0	3.3
		high	.30		

Thus the value of α to be specified in any particular situation depends on the degree of fractionation of the design, or more correctly, the degree of fractionation will depend on the experimenter's expectation of the number of important factors, which would also be reflected in the value of α. For full factorials or half-fractions, a reasonable range for α would be from 0.4 to 0.8, while for more highly fractionated designs, the range would be reduced to 0.2 to 0.4.

In all but one example the value of k_1 for main effects is larger than k_2 for two-factor interactions, and the ratio of the average k_1 to the average k_2 is 3.33.

In practice it will often be informative to carry out the analysis under differing sets of assumptions, e.g., assuming higher order interactions inert or not, or trying different values of α and k. When results are insensitive to varying *plausible* assumptions, one can feel safe in drawing inferences from those results. On the other hand, when the results are not robust to changes in assumptions, this indicates an inability of the data to dominate the information provided initially, and any conclusions should reflect this dependence on prior assumptions. This is illustrated by example in the next section.


AD-A160 972 ANALYSIS OF FACTORIAL EXPERIMENTS(U) WISCONSIN 2/2
UUNIV-MADISON MATHEMATICS RESEARCH CENTER R D MEYER
SEP 85 MRC-TSR-2865 DAAG29-80-C-0041
UNCLASSIFIED F/G 12/1 NL
END



3.4. Example

The method is illustrated by continuing the analysis of Example 2.1. The factors and their allocation to the 16×16 design array are given in Table 3.2. Recall that in Chapter 2 it was discovered that column contrasts 4, 12 and 13 were very likely active, each receiving posterior probability close to one. There was also weak evidence to suggest contrast 8 might also be active (see Figure 2.2a). Assuming interactions between three or more factors to be inert, contrasts 8, 12 and 13 are associated with the main effects of factors 1 (screw speed), 5 (holding pressure) and 6 (booster pressure). The large contrast of column 4 is associated with the sum of four two-factor interactions denoted by the alias string **15+26+37+48**. The original authors suggested that it was most likely either the **15** or **26** interaction which was responsible for the large contrast, because these involved variables with large main effects. In a four-run followup experiment they were able to obtain separate estimates of the four interactions and deduced that the **15** interaction was indeed the major component of the aliased contrast.

The posterior probabilities that each factor is active were computed with α=0.3, k_1=11 and k_2=3.3 and are presented in Table 3.3, again assuming that interactions between three or more factors are inert. Factors 1 (screw speed), 5 (holding pressure) and 6 (booster pressure) have posterior probabilities close to one and could plausibly be considered active. Factor 2 (temperature) has posterior probability of 0.4, with all other factors receiving very small values. Examination of the alias strings of Table 3.3 suggests where the evidence for factor 2 is coming from. Although it does not

Table 3.2 Factors and column allocation to the standard 16-run two-level factorial design for Example 2.1.

	Factors
1	(S) Screw speed
2	(T) Temperature
3	(M) Moisture
4	(V) Thickness
5	(H) Holding pressure
6	(B) Booster pressure
7	(C) Cycle time
8	(G) Gate size

Column allocation

									S	T	M	V	H	B	C	G
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	+	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+
2	+	+	-	-	-	-	+	+	-	-	+	+	+	+	-	-
3	+	-	+	-	-	+	-	+	-	+	-	+	+	-	+	-
4	+	+	+	+	-	-	-	-	-	-	-	-	+	+	+	+
5	+	-	-	+	+	-	-	+	-	+	+	-	-	+	+	-
6	+	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+
7	+	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
8	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
9	+	-	-	+	-	+	+	-	+	-	-	+	-	+	+	-
10	+	+	-	-	-	-	+	+	+	+	-	-	-	-	+	+
11	+	-	+	-	-	+	-	+	+	-	+	-	-	+	-	+
12	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-
13	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+
14	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-
15	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 3.3 Posterior probabilities p_i^* of factors being active, Example 2.1, α=0.3, k_1=11, k_2=3.3, interactions between three or more factors assumed inert. Below are the column contrasts and their alias strings.

	Factors	Posterior probability
1	(S) Screw speed	.875
2	(T) Temperature	.400
3	(M) Moisture	.002
4	(V) Thickness	.004
5	(H) Holding pressure	1.000
6	(B) Booster pressure	.998
7	(C) Cycle time	.003
8	(G) Gate size	.009

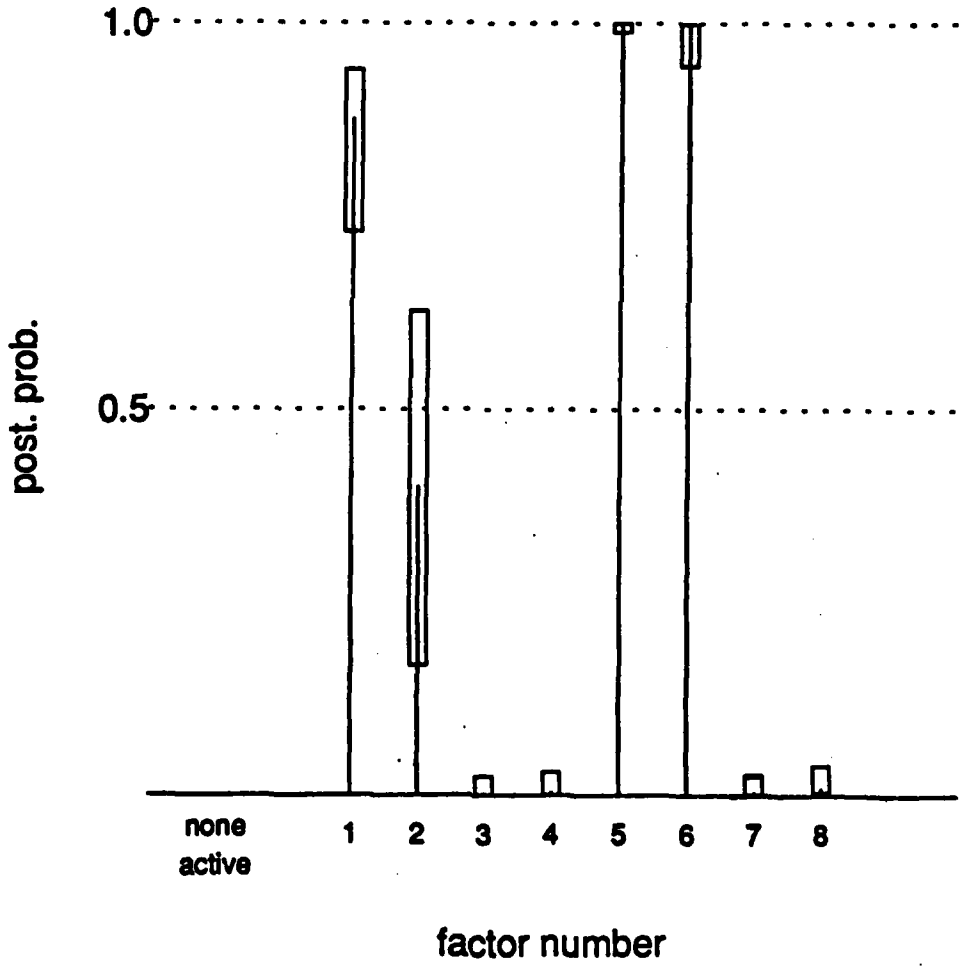
column	contrast	alias string
1	-0.6	12+34+56+78
2	-0.4	13+24+57+68
3	-0.6	14+23+58+67
4	4.6	15+26+37+48
5	0.9	16+25+38+47
6	-0.2	17+28+35+46
7	-0.3	18+27+36+45
8	-1.2	1
9	0.7	2
10	0.1	3
11	0.3	4
12	-5.5	5
13	3.8	6
14	0.1	7
15	-0.6	8

have a large main effect, the two largest interactions could be explained by the effect of variable 2 (as well as variable 1) interacting with the active variables 5 and 6. Variable 1 received higher posterior probability because of its larger main effect.

Alternatively, if variables 1, 5 and 6 were truly the active factors, there is one other variable which would be difficult to separate from those three, and that is variable 2. The design collapses into a full factorial in any combination of four factors which includes variables 1, 5 and 6 except the combination 1, 2, 5 and 6, for which the design collapses into a replicated half fraction. Thus the three two-factor interactions among the variables 1, 5 and 6 are confounded with the three two-factor interactions between 1, 5 and 6 and variable 2. The structure of the design dictates that, given that 1, 5 and 6 are the active factors, it will be more difficult to accumulate evidence against variable 2 than the remaining four factors. This phenomenon is reflected in the results of the Bayesian analysis.

In Figure 3.1 the posterior probabilities are plotted as a bar plot, with boxes indicating the range for each probability over different combinations of α=0.2, 0.3 and 0.4, k_1=5, 11, and 15, and k_2=2, 3.3 and 6, only taking those combinations with $k_1 > k_2$. The posterior probability for factor 2 is the only one which changes enough to affect conclusions about the experiment. Conclusions about this factor depend upon assumptions about the frequency of active factors and the relative size of main effects, interactions and inert contrasts, and these assumptions are reflected in the values of α, k_1 and k_2. In particular, if knowledge of these parameters is vague, variable 2 can not be safely eliminated.

Figure 3.1 Posterior probabilities $\{p_i^*\}$ that factors are active, Example 2.1. Solid lines indicate values for α=0.3, k_1=11, k_2=3.3, boxes indicate range of values for different combinations of α=0.2, 0.3, 0.4, k_1=5, 11, 15, and k_2=2, 3.3, 6.



Suppose now that the assumption that three-factor interactions are inert is dropped. Although this is a very reasonable assumption in practice, it will be interesting to observe what occurs when it is dropped. The posterior probabilities of the factors were recomputed based on the new set of assumptions and are presented in Table 3.4. The evidence for variables 5 and 6 is still strong, but the posterior probabilities for variables 1 and 2 are now almost equal. The reason can be found in the revised alias strings for each contrast in Table 3.4. The contrast associated with the main effect of variable 1 is confounded with the 256 interaction. Now that this contrast, which is a bit too large to attribute entirely to noise, can be associated with an effect of variables 2, 5 and 6, the evidence for variable 2 is stronger, and the evidence for variable 1 is somewhat weaker. This alternative analysis was presented to demonstrate that an experimenter who might have eliminated variable 2 based on the previous analysis would have depended heavily on the assumption that three-factor interactions were inert.

There are two separate issues to consider when making assumptions: the reasonability of the assumptions, and the dependence of the conclusions on the assumptions. It was shown for Example 2.1 that conclusions could be sensitive to choice of prior parameters and the assumption that three-factor interactions are inert. While the reasonability of such assumptions is not questioned, it is important to know when conclusions depend on assumptions even when the assumptions are well-based.

To demonstrate further the point made about confounding when there are R-1 active factors for a design of resolution R, consider the following exercise. Suppose

Table 3.4 Posterior probabilities of factors being active, Example 2.1, α=0.3, k_1=11, k_2=3.3, interactions between four or more factors assumed inert. Below are the column contrasts and their alias strings (only three-factor interactions among the plausibly active factors 1, 2, 5 and 6 are shown).

	Factors	Posterior probability
1	(S) Screw speed	.608
2	(T) Temperature	.537
3	(M) Moisture	.000
4	(V) Thickness	.000
5	(H) Holding pressure	.991
6	(B) Booster pressure	.942
7	(C) Cycle time	.000
8	(G) Gate size	.000

column	contrast	alias string
1	-0.6	12+34+56+78
2	-0.4	13+24+57+68
3	-0.6	14+23+58+67
4	4.6	15+26+37+48
5	0.9	16+25+38+47
6	-0.2	17+28+35+46
7	-0.3	18+27+36+45
8	-1.2	1+256
9	0.7	2+156
10	0.1	3
11	0.3	4
12	-5.5	5+126
13	3.8	6+125
14	0.1	7
15	-0.6	8

the three factors 1, 5 and 6 were the only active factors. Their activity could be manifested in several different combinations of main effects and interactions. Assuming the 156 interaction is inert, that leaves three main effects and three two-factor interactions among the active factors. For purposes of illustration, artificial data will be created to explore the relationships among the factors 1, 2, 5 and 6 under these circumstances. Suppose main effects are always either 2 or 0, and two-factor interactions are either 1 or 0. Since each effect can take on either of two values, there are 2^6=64 possible combinations of main effects and interactions. For each combination, a vector of observations y is generated, with no error component, and the posterior probabilities $\{p_i^*\}$ are computed, with α=0.3, k_1=11 and k_2=3.3.

Of the 64 possible combinations of the six effects, 23 correspond to situations when not all three factors are active, for example when all six effects are zero. These are eliminated from further consideration. The remaining 64-23=41 can be represented by 12 distinct combinations. For example, there are three ways to have three non-zero main effects and one non-zero interaction, but each of these gives the same pattern of values for the posterior probabilities. The 12 distinct combinations and the probabilities $\{p_i^*\}$ for factors 1, 2, 5 and 6 are presented in Table 3.5. (The remaining factors received posterior probability of zero, to two decimal places, for all 12 combinations).

As seen in the table, there are many situations in which factor 2 receives significant posterior probability, despite the fact it is actually inert and there is no error component in the data. The combinations for which it was easiest to detect the

Table 3.5 Posterior probabilities of factors being active over the 12 distinct combinations of active effects for the active factors 1, 5 and 6, α=0.3, k_1=11, k_2=3.3. The first six columns give the values assigned to the main effects and interactions among factors 1, 5 and 6, and the last four columns give the posterior probabilties for factors 1, 2, 5 and 6.

Effects						Posterior Probabilities			
$b1$	$b5$	$b6$	$b15$	$b16$	$b56$	1	2	5	6
2	2	2	0	0	0	1.00	.01	1.00	1.00
2	2	2	1	0	0	1.00	.09	1.00	1.00
2	2	2	1	1	0	1.00	.21	1.00	1.00
2	2	2	1	1	1	1.00	.30	1.00	1.00
2	2	0	0	1	0	1.00	.54	1.00	.54
2	2	0	1	1	0	1.00	.59	1.00	.59
2	2	0	0	1	1	1.00	.59	1.00	.59
2	2	0	1	1	1	1.00	.62	1.00	.62
2	0	0	1	1	0	1.00	.74	.74	.74
2	0	0	1	0	1	1.00	.74	.74	.74
2	0	0	1	1	1	1.00	.76	.76	.76
2	0	0	0	0	1	1.00	.99	.01	.01

truly active factors were those in which all main effects were large. As main effects were dropped it became more difficult to separate factor 2 from the other three factors. This is because the assumption that main effects are larger and occur more frequently than interactions has been incorporated into the model, and situations for which this does not hold can lead to these unexpected patterns of probabilities. However, on the premise that the assumption about main effects is basically sound, these troublesome situations should not be frequent. Note also there was only one combination where active factors did not receive large probabilities, and this was the combination where there was a nonzero interaction between factors 5 and 6, but their respective main effects were zero.

Two conclusions are apparent from this exercise. First, when there are $R-1$ active factors in a design of resolution R, it is sometimes not possible to identify the correct factors exactly, even though the design projects into a full factorial in the $R-1$ factors. Fortunately, it will usually be possible to restrict attention to some subset of variables, and it would be rare that active factors would be excluded from this subset due to inherent properties of the design and analysis (active factors may be concealed by noise). In the example above it was possible to narrow down to four of the original eight variables. A follow-up experiment such as the one described in Box, Hunter and Hunter (1978), p. 413, can be implemented to eliminate any remaining inert factors. Second, the Bayesian analysis provides a good method for identifying the likely subset of variables by combining prior assumptions, properties of the experimental design and information in the data. Factors such as factor 2 in the above example

which cannot be safely eliminated because of the structure of the design, are identified by their non-negligible posterior probability, as well as those factors which are more obviously active.

The above exercise was repeated with pseudo-random normal errors added to the artificially generated observations, 100 trials for each combination. The average posterior probabilities achieved over the 100 trials agreed closely with those in Table 3.5.

3.5. Robustness and the Assumption of Normal Errors

Sensitivity of the Bayesian analysis to the assumption of normal errors is explored in this section. As described by Box and Tiao (1973), the idea of robustness has two facets, criterion robustness and inference robustness. Criterion robustness is concerned with the performance of a statistical procedure derived from one set of assumptions when a different set of assumptions is true. A procedure, or criterion, would be robust if it performed similarly under both sets of conditions. For example, when estimating the mean of a population, normal theory confidence limits around the sample mean may still apply approximately when the data are non-normal, because of the central limit theorem. Inference robustness is concerned with the comparison of procedures derived from different sets of assumptions. If the procedure derived from assumptions A_1 leads to nearly the same inference as the procedure derived from assumptions A_2, then that inference would be robust. For example, measuring the sensitivity of the posterior probabilities to choice of α and k, Section 2.3.2, dealt with the question of inference robustness. As a general principle one would only be concerned about robustness over reasonably plausible assumptions.

To assess the robustness of inferences from the posterior probabilities $\{p_i\}$ or $\{p_i^*\}$ with respect to choice of error distribution, it would be necessary to derive new formulas for these probabilities based on some non-normal distribution. However, working with reasonable alternative distributions such as the t, double exponential, rectangular, etc., the numerous integrations in the expression for the posterior probabilities could not be handled analytically. Multi-dimensional numerical integration

also proved to be intractable, and this issue is left for future consideration.

The issue of criterion robustness to the assumption of normality is more straightforward. In the following sections the results of simulations designed to explore this issue are described.

3.5.1. No Active Effects

For the first set of simulations data were generated with no active effects present. Behavior of the posterior probabilities $\{p_i\}$ and $\{p_i^*\}$ was observed over four error distributions: normal, rectangular (a platykurtic or light-tailed distribution), t with 3 degrees of freedom, abbreviated by t_3 (a leptokurtic or heavy-tailed distribution), and skew normal (errors were generated from a normal distribution with zero mean, and positive values were multiplied by a constant greater than one to create a skewed distribution; the constant was chosen to give a coefficient of skewness of 1, equal to the skewness of a chi-square random variable with 8 degrees of freedom, for example). For each of these distributions, 100 pseudo-random samples of size n=16 and standard deviation 1.0 were generated, and for each sample the $n-1$ othogonal contrasts were obtained from the design array in the usual way. From these, the posterior probabilities $\{p_i\}$ and $\{p_i^*\}$ were computed for each sample, assuming a 2^{8-4} two-level design was carried out, with α=0.2 and k=10 for computing $\{p_i\}$ and α=0.3, k_1=11 and k_2=3.3 for computing $\{p_i^*\}$.

For each error distribution the 100 sets of posterior probabilities were summarized and plotted in the following way. Because there were no real effects, the variation in the probabilities for each column (or factor) are roughly the same. Thus it is

more informative to order the probabilities for each sample and then examine the behavior of these, and the probability of no active columns (or factors), over the 100 generated samples. For example, the location and variation of the maximum probability from each sample in relation to the probability of no active columns is of primary interest. In Figures 3.3 and 3.4 the median probability for each category (none active, largest probability, second largest probability, etc.) is plotted as an asterisk. A box around the asterisk represents the inter-quartile range over the simulations. Vertical lines from the ends of this box extend to the upper and lower fifth percentiles (5%, 95%). This is illustrated in Figure 3.2.

On the left in each of Figures 3.3a-d the posterior probability of no active columns is plotted. Since the data were generated with no real effects, this value is expected to be larger than the other probabilities. Although it is apparent from Figure 3.3 that this does not always occur, i.e., the largest column posterior probability is often larger than the probability of no active columns, this is partly due to the low *prior* probability of no active columns, which is $(1-.2)^{15} = .035$. Likewise the prior probability that the largest contrast is active is something in the neighborhood of one minus this probability, or .965. Thus it is difficult with only 16 observations to reverse these probabilities. It is encouraging just the same that 75% of the time the largest column posterior probability is still less than 1/2 over all four distributions.

Comparing the patterns of variation in the posterior probabilities across the four error distributions, they are in excellent agreement. The only notable deviation is the smaller column posterior probabilities when the t_3 and skew normal distributions

Figure 3.2 Plotting convention for Figures 3.3-3.6. The asterisk indicates the median value over the simulations, the box extends from the lower to the upper quartile, and vertical lines extend to the upper and lower fifth percentiles.

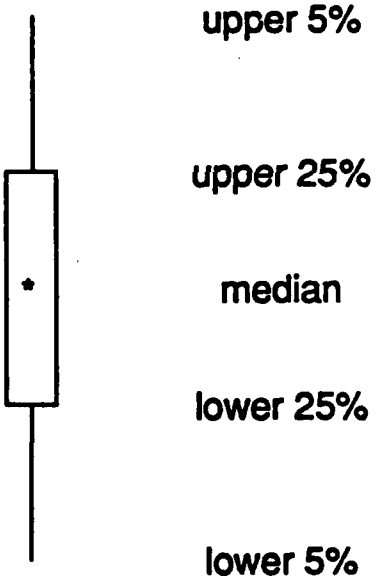


Figure 3.3a Plot of posterior probabilities $\{p_i\}$ over 100 simulations, normal errors, no real effects, α=0.2, k=10. The column probabilities $p_1,\ldots,p_{15}$ were ordered for each simulated sample so that, for example, the label 1 on the x-axis refers to the probability associated with the largest contrast. See Figure 3.2 for explanation of plotting symbols.

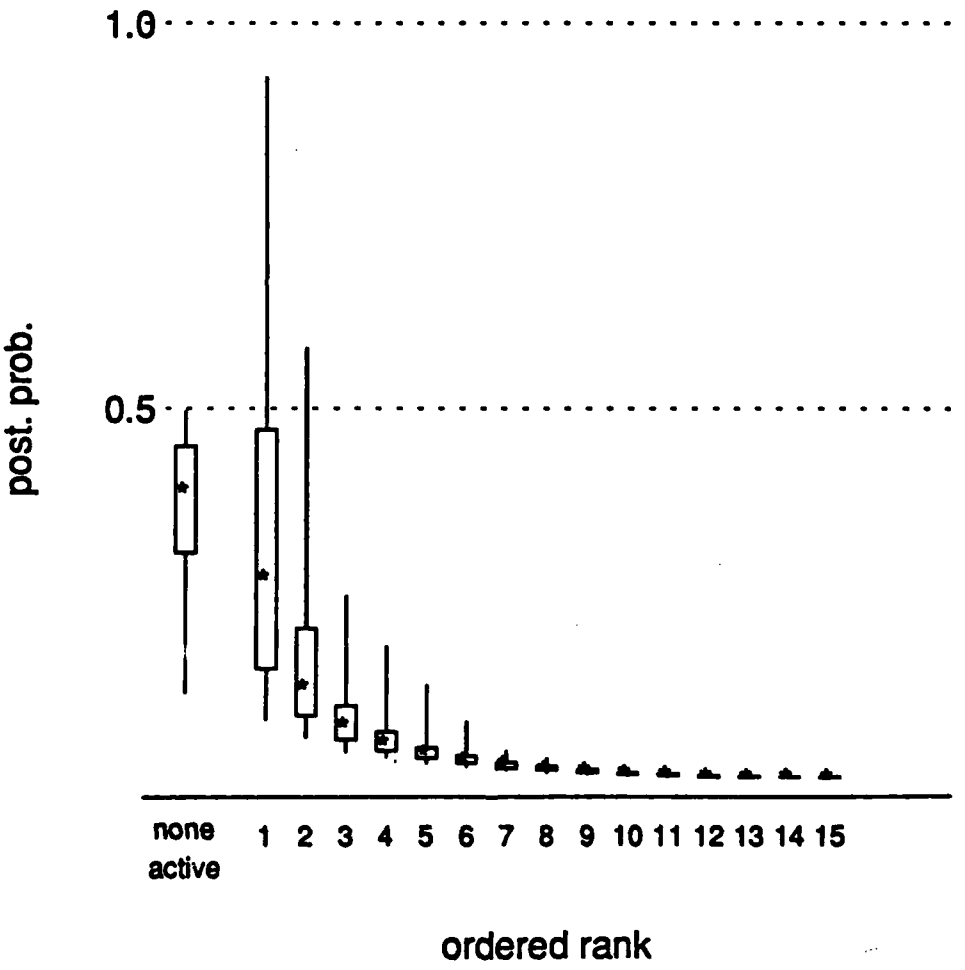


Figure 3.3b Plot of posterior probabilities $\{p_i\}$ over 100 simulations, rectangular errors, no real effects, α=0.2, k=10. See Figures 3.2, 3.3a for plotting conventions.

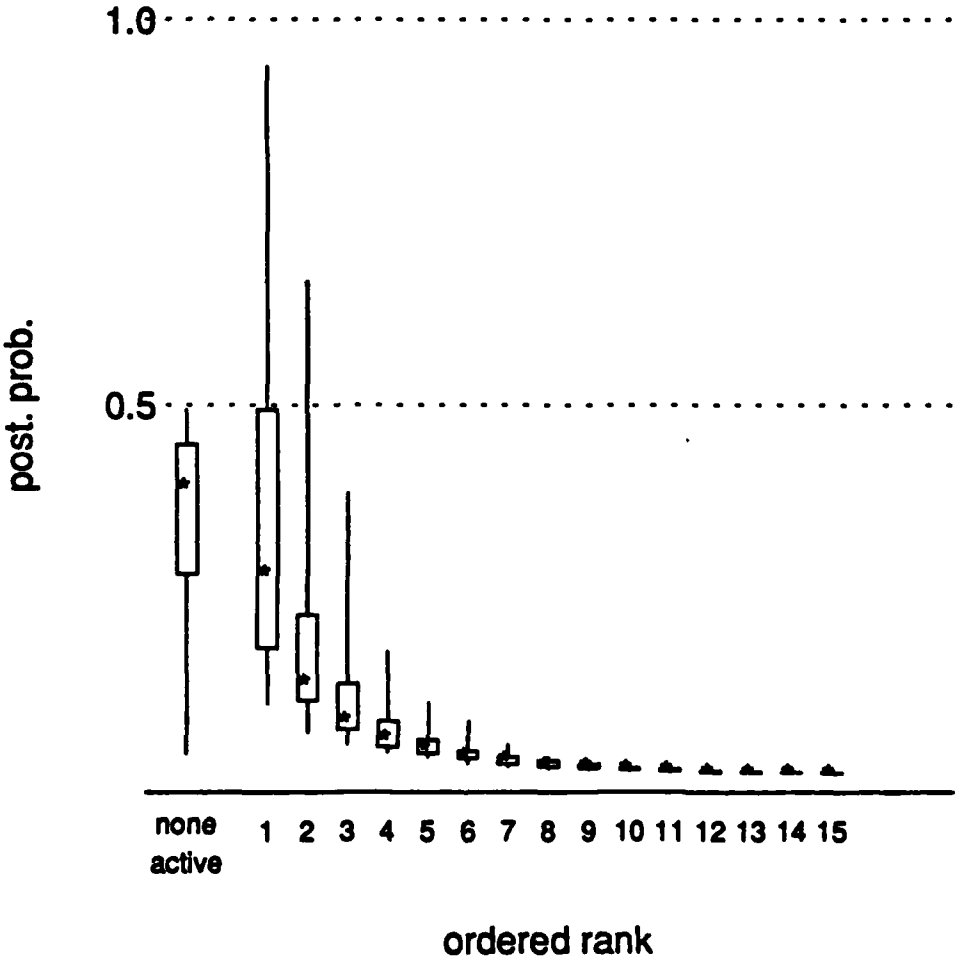


Figure 3.3c Plot of posterior probabilities $\{p_i\}$ over 100 simulations, t_3 errors, no real effects, α=0.2, k=10. See Figures 3.2, 3.3a for plotting conventions.

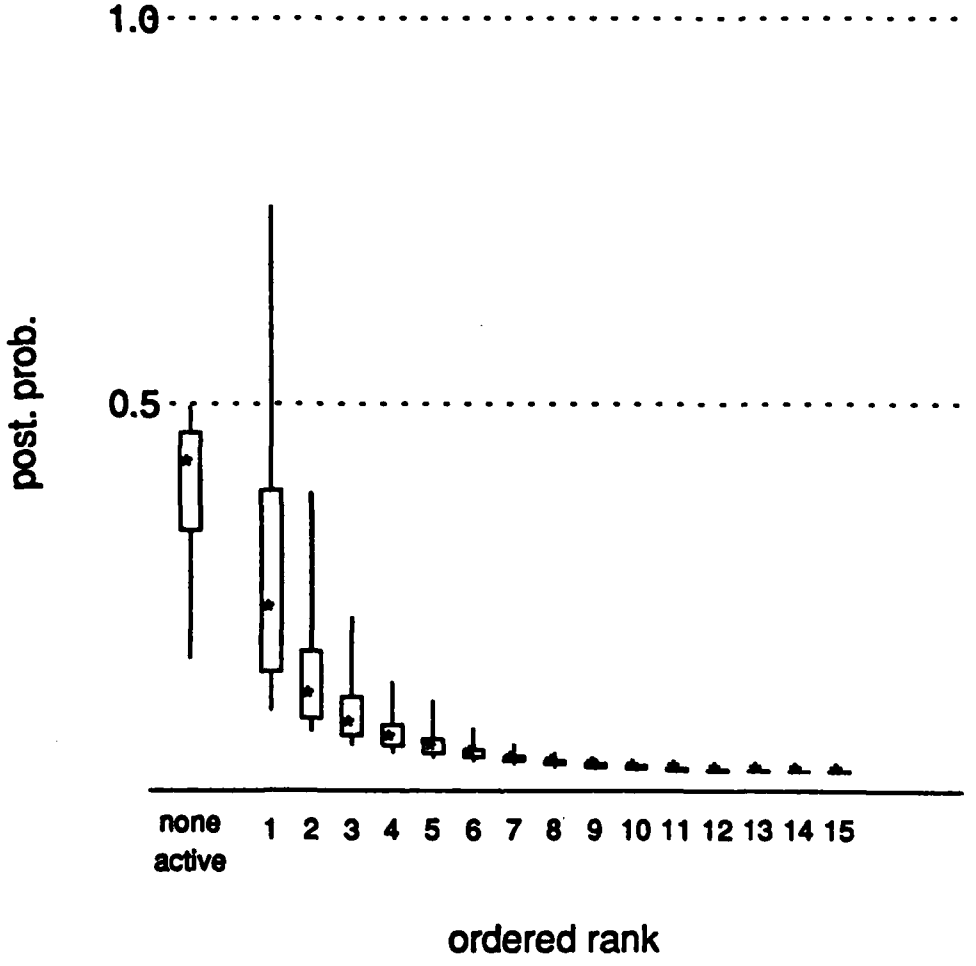
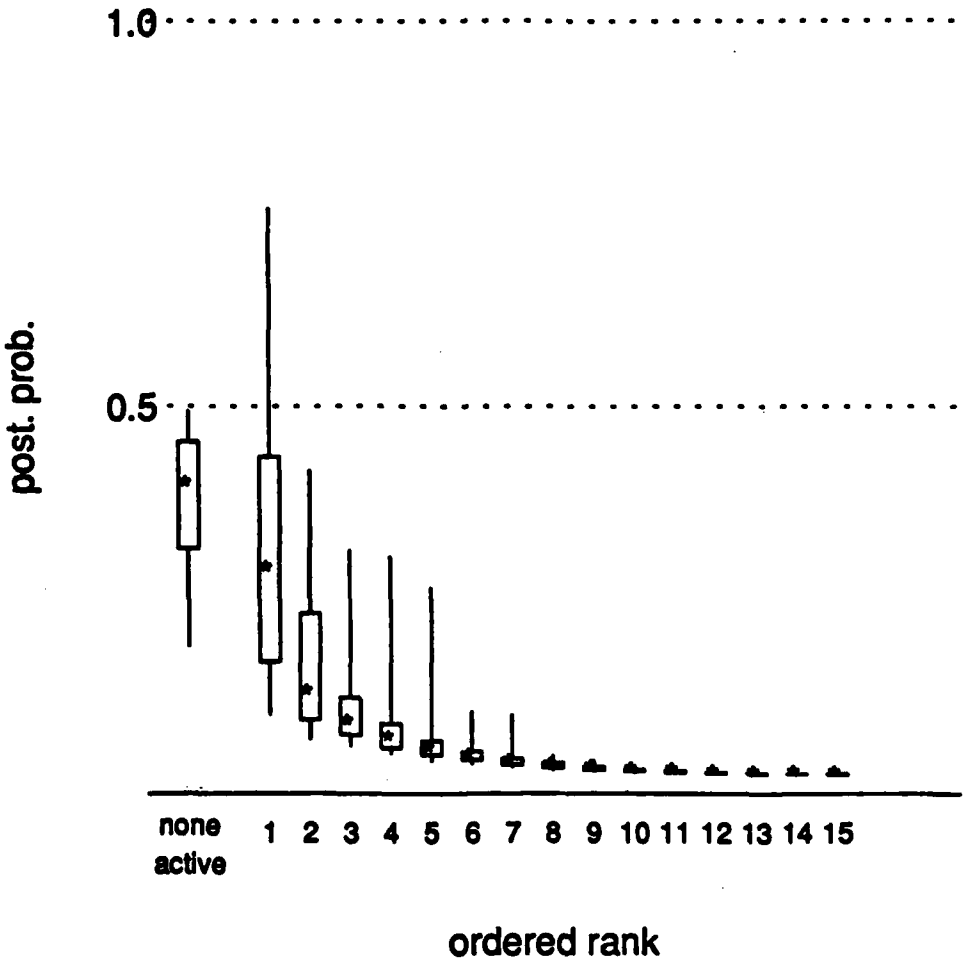


Figure 3.3d Plot of posterior probabilities $\{p_i\}$ over 100 simulations, skew normal errors, no real effects, α=0.2, k=10. See Figures 3.2, 3.3a for plotting conventions.



were used. This difference from the observed pattern for normal errors is small and in the direction of being more correct, i.e., a smaller posterior probability of being active when the contrast is indeed not active. For comparison, the numerical values of the percentiles of the posterior probability of no active columns and the maximum column probability are given in Table 3.6.

The behavior of the factor posterior probabilities $\{p_i^*\}$, Figures 3.4a-d, was very similar to that observed for the column posterior probabilities $\{p_i\}$. The probability $p_{(0)}^*$ of no active factors was larger than 1/2 almost 50% of the time for all four distributions. The overall patterns of variation of $p_{(0)}^*$ over the different distributions agree very closely. The largest factor posterior probability is often larger then $p_{(0)}^*$, but by the same argument as for the column probabilities, this is neither surprising nor alarming.

The patterns of variation in the factor posterior probabilities for the rectangular and t_3 distributions differed somewhat from that observed for the normal case. Again, the probabilities for the t_3 distribution were lower, but deviations in that direction are not troublesome. For the rectangular distribution, there was a tendency to get slightly higher probabilities, although, for example, the median maximum probability is the same as for the normal. The fear is that the Bayesian analysis will find active contrasts when there are none, because the errors are not normal. The difference observed for the rectangular error distribution is not large enough to validate that fear. The numerical values of the percentiles of the posterior probability of no active factors and the maximum factor probability are given in Table 3.7.

Table 3.6 Percentiles of the distribution of a) the probability of no active columns and b) the maximum column probability, over 100 pseudo-random error samples of size $n = 16$ from the normal, rectangular, t_3 and skew normal distributions, added to a data vector y of zeroes (no real effects present). Probabilities were computed with $\alpha = 0.2$, and $k = 10$.

Probability of None Active				
Percent	Normal	Rectangular	t_3	Skew Normal
95%	.496	.494	.496	.493
75%	.451	.448	.462	.454
50%	.393	.389	.422	.398
25%	.314	.280	.336	.315
5%	.135	.050	.171	.190

Maximum Column Probability				
Percent	Normal	Rectangular	t_3	Skew Normal
95%	.929	.939	.755	.756
75%	.472	.494	.388	.434
50%	.281	.310	.233	.288
25%	.165	.185	.153	.170
5%	.100	.114	.104	.102

Figure 3.4a Plot of posterior probabilities $\{p_i^*\}$ over 100 simulations, normal errors, no real effects, α=0.3, k_1=11, k_2=3.3. The probabilities $p_1^*,\ldots,p_8^*$ were ordered so that, for example, the label 1 on the x-axis refers to the largest probability from each sample, etc. See Figure 3.2 for explanation of plotting symbols.

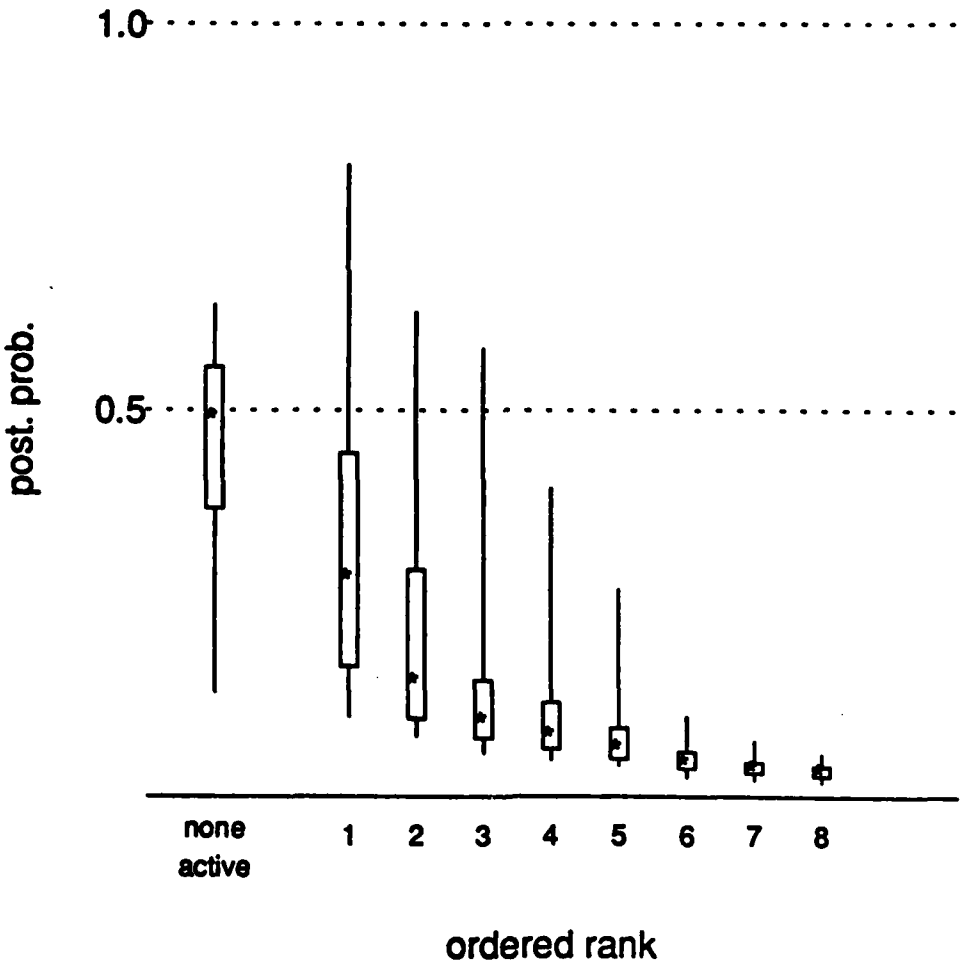


Figure 3.4b Plot of posterior probabilities $\{p_i^*\}$ over 100 simulations, rectangular errors, no real effects, α=0.3, k_1=11, k_2=3.3. See Figures 3.2, 3.4a for plotting conventions.

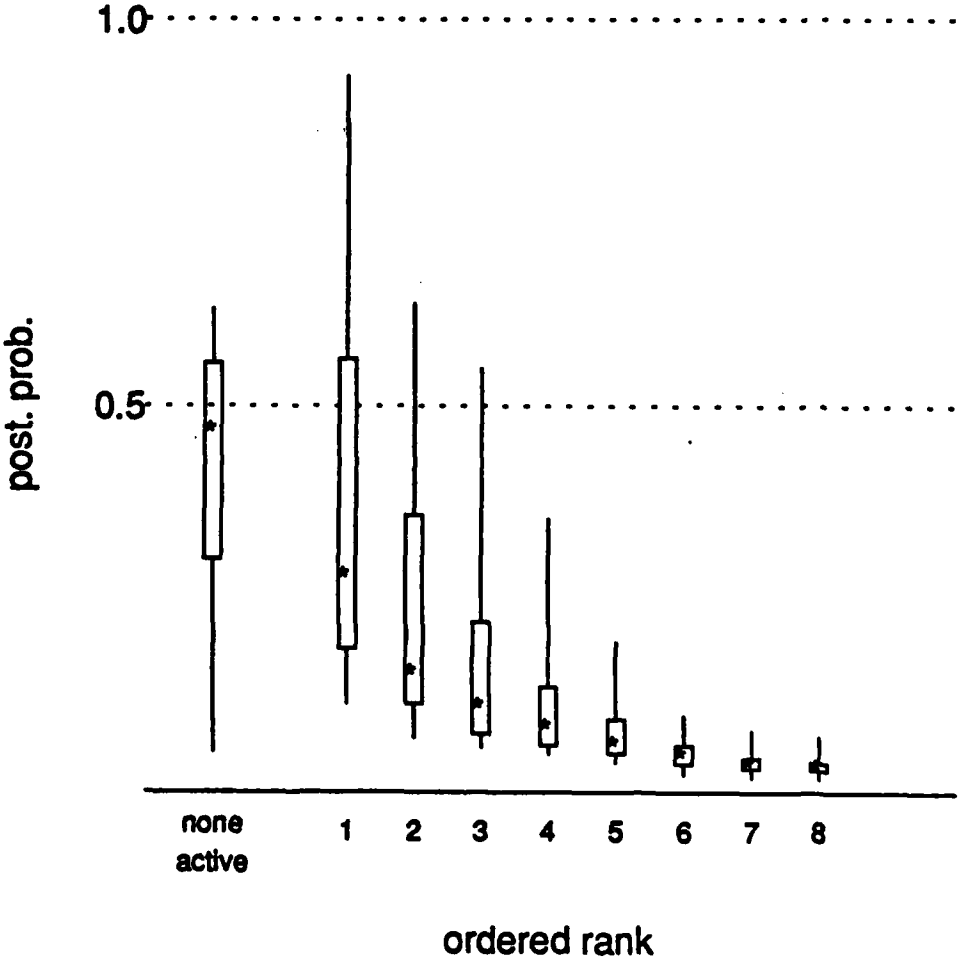


Figure 3.4c Plot of posterior probabilities $\{p_i^*\}$ over 100 simulations, t_3 errors, no real effects, α=0.3, k_1=11, k_2=3.3. See Figures 3.2, 3.4a for plotting conventions.

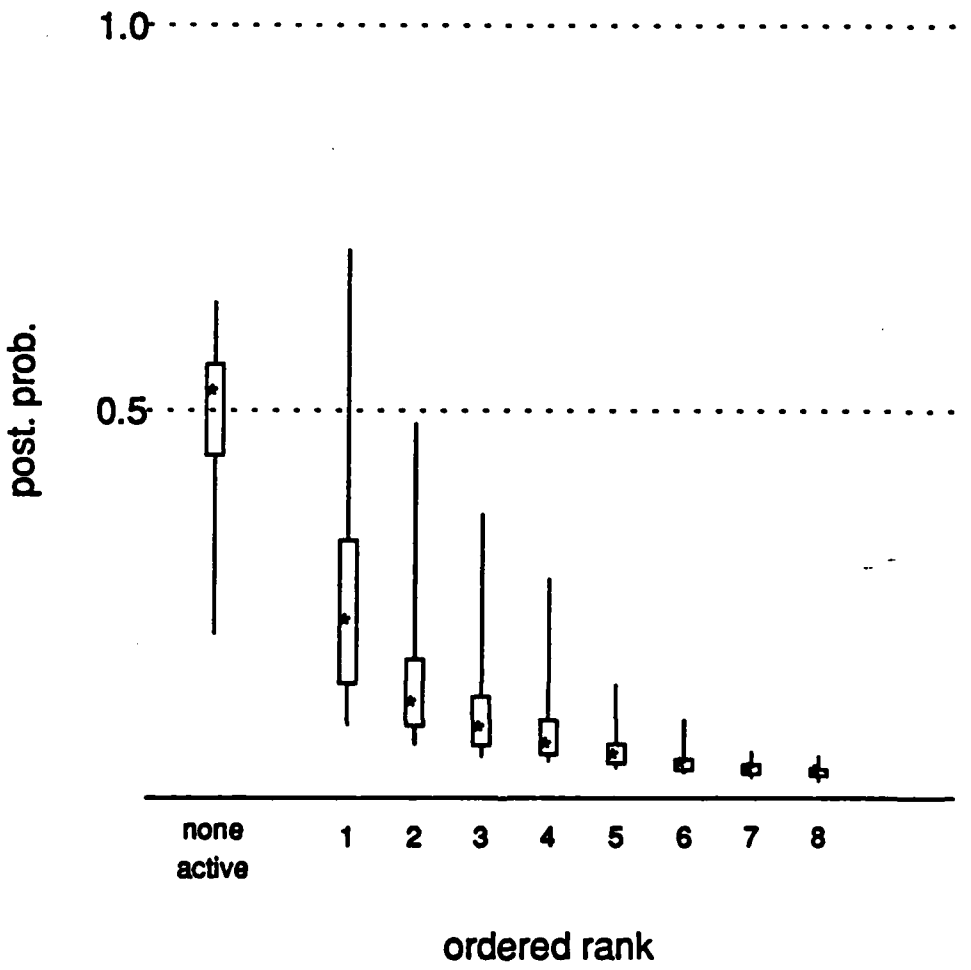


Figure 3.4d Plot of posterior probabilities $\{p_i^*\}$ over 100 simulations, skew normal errors, no real effects, α=0.3, k_1=11, k_2=3.3. See Figures 3.2, 3.4a for plotting conventions.

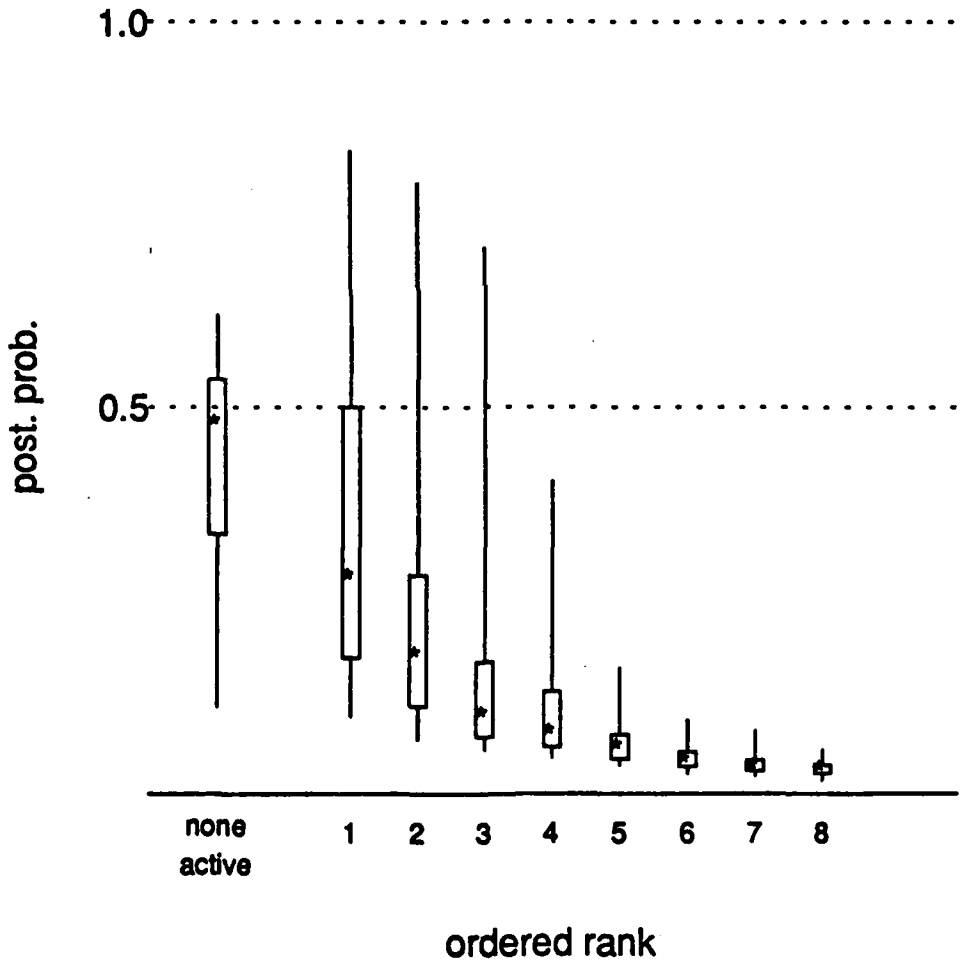


Table 3.7 Percentiles of the distribution of a) the probability of no active factors and b) the maximum factor probability, over 100 pseudo-random error samples of size n = 16 from the normal, rectangular, t_3 and skew normal distributions, added to a data vector y of zeroes (no real effects present). Probabilities were computed pretending a 2^{8-4} design was carried out, with $\alpha = 0.3$, $k_1 = 11$, and $k_2 = 3.3$.

Probability of None Active

Percent	Normal	Rectangular	t_3	Skew Normal
95%	.636	.625	.639	.619
75%	.556	.555	.560	.536
50%	.491	.468	.522	.479
25%	.373	.302	.443	.335
5%	.135	.052	.213	.113

Maximum Factor Probability

Percent	Normal	Rectangular	t_3	Skew Normal
95%	.817	.926	.707	.832
75%	.444	.561	.333	.500
50%	.283	.281	.227	.279
25%	.168	.186	.148	.175
5%	.103	.115	.095	.100

When there are no real effects present, non-normality of the errors does not appear to significantly bias the posterior probabilities. As the frequency of having no real effects may tend to be low in practice, another set of simulations was carried out for the case where there are real effects present.

3.5.2. Five Active Contrasts

The second set of simulations is concerned with the behavior of the posterior probabilities under different error distributions when there are real effects present. A standard observation vector y was generated by taking the five largest contrasts from Example 2.1 corresponding to columns 3, 5, 8, 10 and 12, and setting the remaining contrasts equal to zero. Thus y is the vector of 16 observations which would give the five non-zero contrasts mentioned above and ten contrasts exactly equal to zero. For each of the four error distributions, normal, rectangular, t_3 and skew normal, 100 samples of pseudo-random errors were generated with mean zero (median zero for the skew normal) and second moment of 1.0. The value of 1.0 was chosen because it was close to the estimated variance for the real data of Example 2.1. Each individual error vector was added to the vector y, and the posterior probabilities $\{p_i\}$ and $\{p_i^*\}$ were computed, again assuming a 2^{8-4} design was carried out, with α=0.2 and k=10 for computing $\{p_i\}$ and α=0.3, k_1=11 and k_2=3.3 for computing the $\{p_i^*\}$.

For each error distribution the 100 sets of posterior probabilities were plotted with the same plotting conventions described in Figure 3.2. Because there were real effects present, the probabilities weren't ordered as in the previous set of simulations. The probabilities $\{p_i\}$ are plotted in Figure 3.5a-d, and the probabilities $\{p_i^*\}$ are

plotted in Figure 3.6a-d.

The posterior probability $p_{(0)}$ of no active contrasts as well as the probability $p_{(0)}^{*}$ of no active factors remained essentially zero over all 100 simulations for each of the four error distributions. Non-normal errors do not tend to inflate the probability of none active when real effects are present.

Now consider the probabilities $\{p_i\}$ of Figure 3.5. The probabilities p_3, p_5 and p_{12} corresponding to the three very large contrasts remain uniformly close to one over the 100 simulations and over all four distributions. Similarly the probabilities corresponding to the ten inert contrasts tended to be fairly close to zero, and exhibited the same sort of variation for each distribution. One could safely conclude that non-normal errors would not lead to a gross error of judging an unquestionably active effect as inert or an obviously inert one as active.

The probabilities p_8 and p_{10} corresponding to the marginally active contrasts of columns 8 and 10 exhibit the most variation over the 100 simulations, and the patterns of variation among the four distributions are different. The skew normal case agreed quite closely with the normal case, while the rectangular and t_3 distributions tended to give larger values to probabilities p_8 and p_{10}. The implications of this are not particularly worrisome. First of all, the preferred error in most instances would be to mistake an inert contrast for an active one, rather than missing a real effect. Larger posterior probabilities for marginal contrasts would have that effect. Secondly, contrasts with probabilities in the interval (0.2,0.8) would generally be judged to need further investigation before very firm conclusions could be made. If a probability of

Figure 3.5a Plot of posterior probabilities $\{p_i\}$ over 100 simulations, normal errors, σ=1, α=0.2, k=10, columns 3, 5, 8, 10 and 12 have real effects equal to those from Example 2.1. The remaining columns were assigned zero effects. See Figure 3.2 for explanation of plotting symbols.

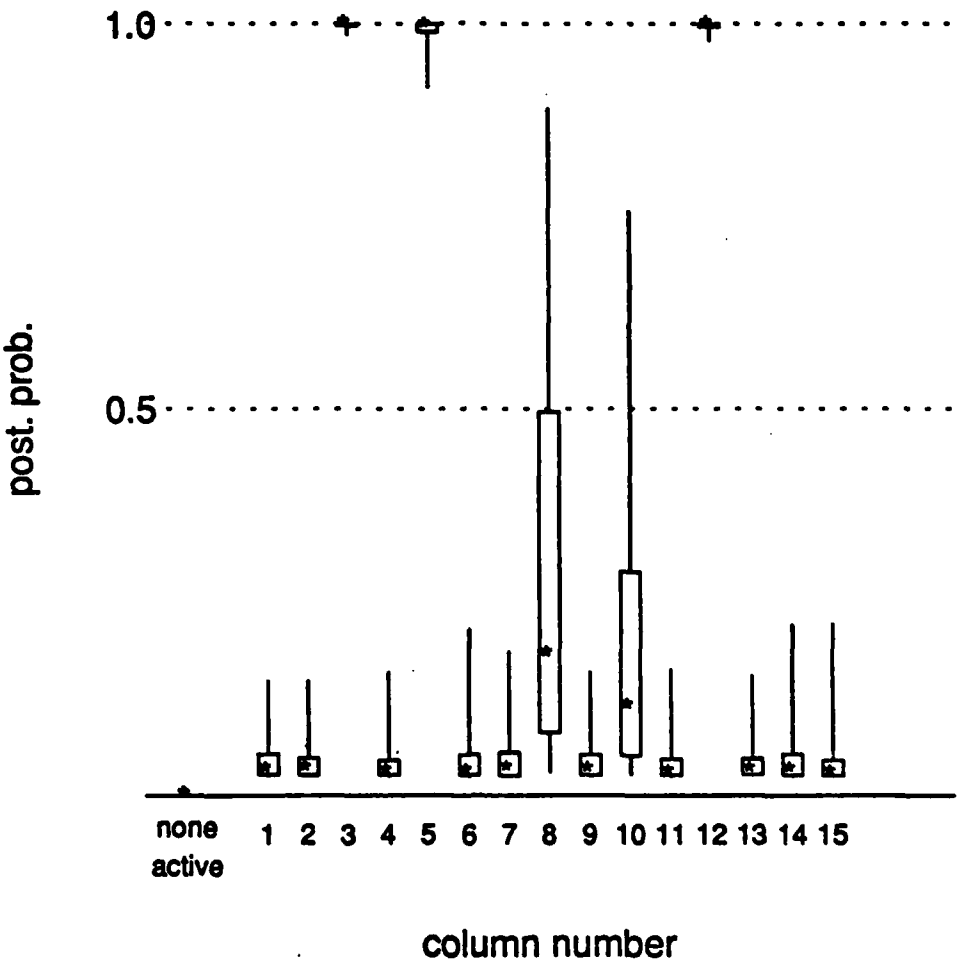


Figure 3.5b Plot of posterior probabilities $\{p_i\}$ over 100 simulations, rectangular errors, σ=1, α=0.2, k=10, columns 3, 5, 8, 10 and 12 have real effects equal to those from Example 2.1. The remaining columns were assigned zero effects. See Figure 3.2 for explanation of plotting symbols.

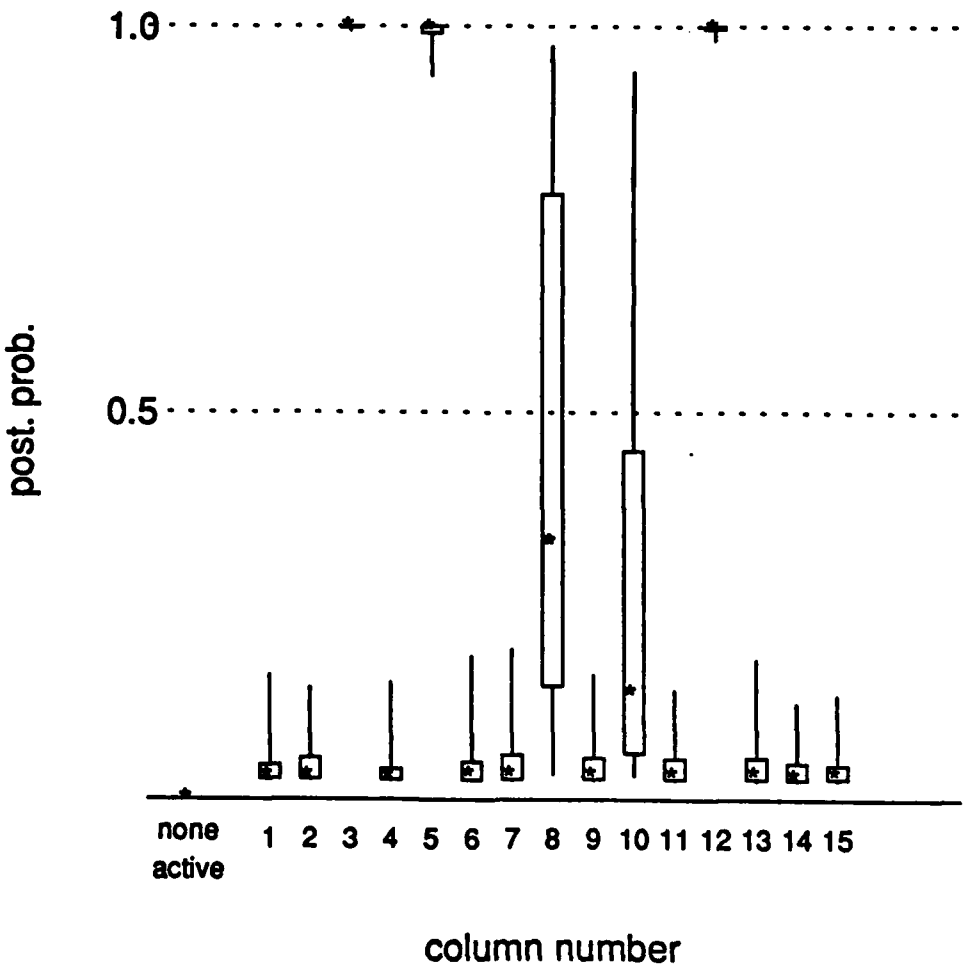


Figure 3.5c Plot of posterior probabilities $\{p_i\}$ over 100 simulations, t_3 errors, σ=1, α=0.2, k=10, columns 3, 5, 8, 10 and 12 have real effects equal to those from Example 2.1. The remaining columns were assigned zero effects. See Figure 3.2 for explanation of plotting symbols.

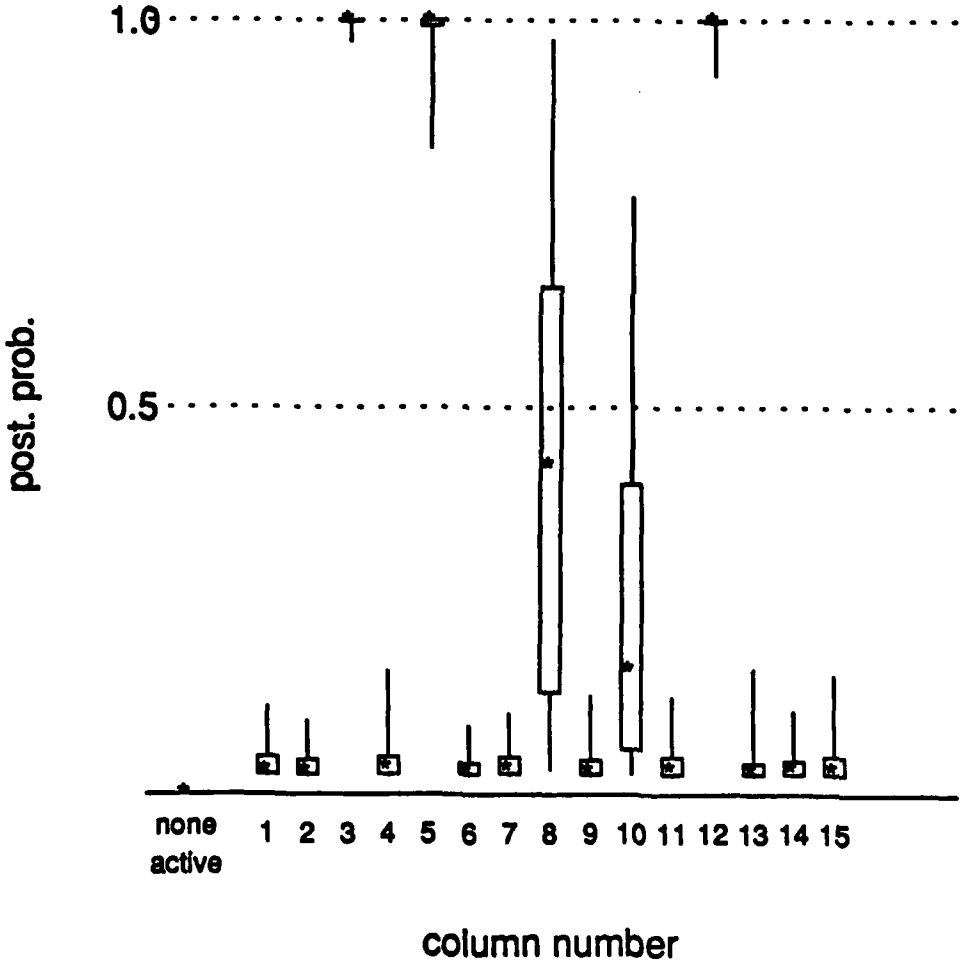
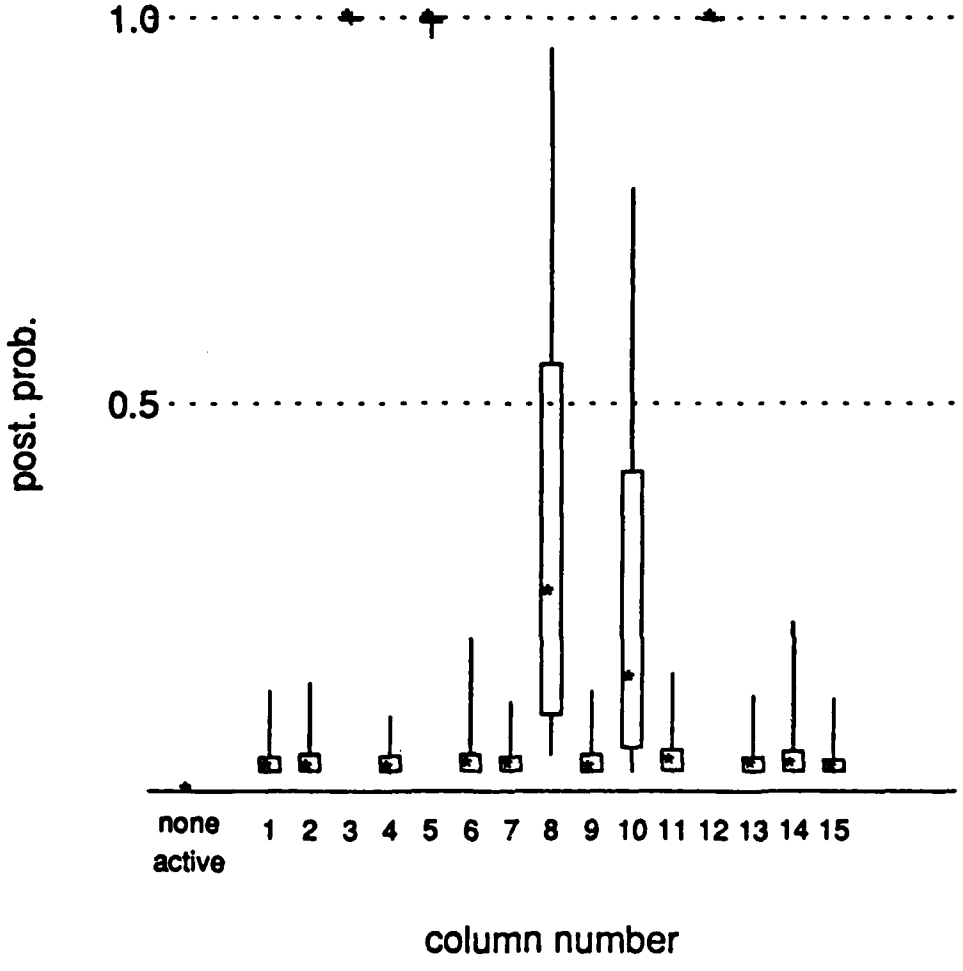


Figure 3.5d Plot of posterior probabilities $\{p_i\}$ over 100 simulations, skew normal errors, σ=1, α=0.2, k=10, columns 3, 5, 8, 10 and 12 have real effects equal to those from Example 2.1. The remaining columns were assigned zero effects. See Figure 3.2 for explanation of plotting symbols.



0.6 were observed rather than 0.4, the investigator's suspicions about that contrast would be essentially the same.

Considering now the probabilities $\{p_i^*\}$ of Figure 3.6, the patterns of variation in these are in close agreement for all four distributions. The probabilities corresponding to factors 3, 4, 7 and 8 are uniformly close to zero for each distribution, and the probabilities for factors 5 and 6 are uniformly close to one. The most variation was observed in the probabilities associated with the factors 1 and 2, but the patterns of variation were very similar across the four distributions. Thus there is no evidence that non-normal errors would have an adverse effect on the $\{p_i^*\}$.

Overall, the variational pattern of the posterior probabilities $\{p_i\}$ and $\{p_i^*\}$ are quite similar for the four error distributions with and without active columns, validating a claim of criterion robustness to the normality assumption for the Bayesian analysis. At the same time, the results of the simulations also verified that the posterior probabilities lead to sensible inferences, as inert columns and factors consistently received low posterior probability and active columns and factors consistently received high posterior probability.

3.6. Conclusions

The method derived in this chapter provides an interesting new way of analyzing factorial experiments. Its major attraction is how it combines prior assumptions, properties of the design and information in the data to identify active factors. A unified analysis of this sort has never really been available before. Varying assumptions about the size and relative frequency of main effects and interactions can also indicate

Figure 3.6a Plot of posterior probabilities $\{p_i^*\}$ over 100 simulations, normal errors, σ=1, α=0.3, k_1=11, k_2=3.3, five active columns from Example 2.1, factors 1, 2, 5 and 6 possibly active. See Figure 3.2 for explanation of plotting symbols.

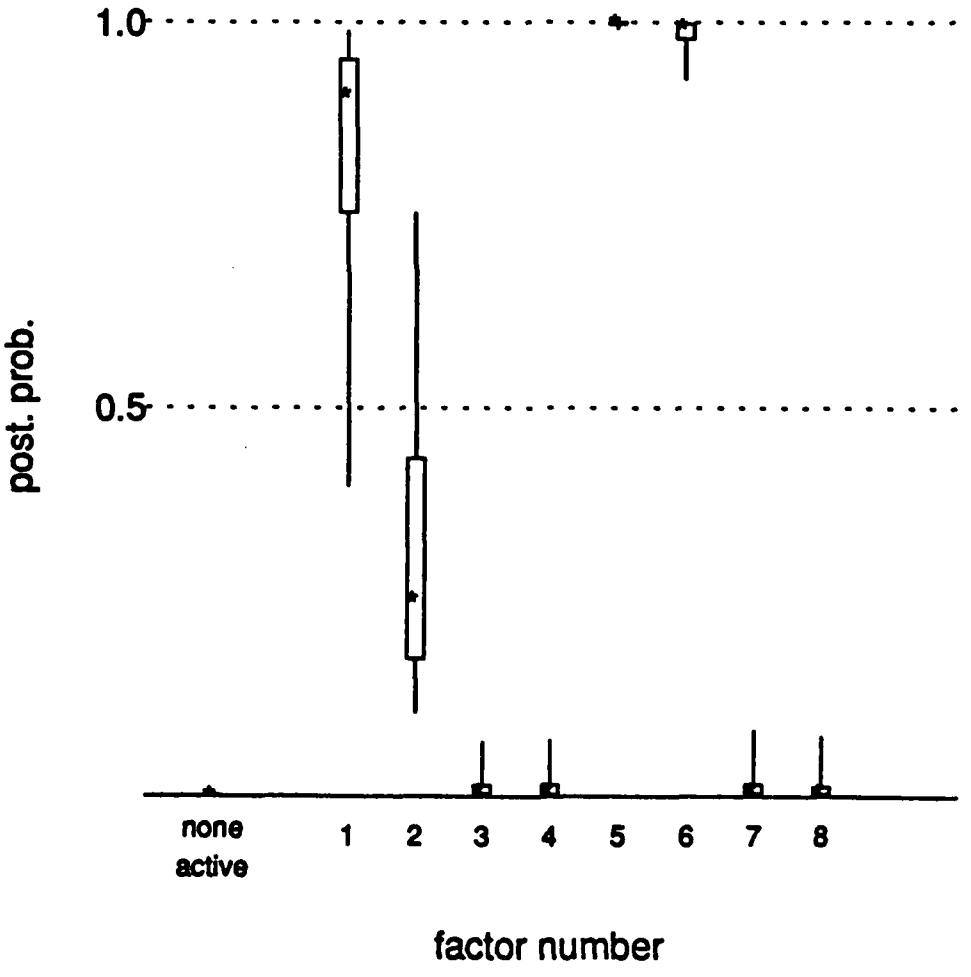


Figure 3.6b Plot of posterior probabilities $\{p_i^*\}$ over 100 simulations, rectangular errors, σ=1, α=0.3, k_1=11, k_2=3.3, five active columns from Example 2.1, factors 1, 2, 5 and 6 possibly active. See Figure 3.2 for explanation of plotting symbols.

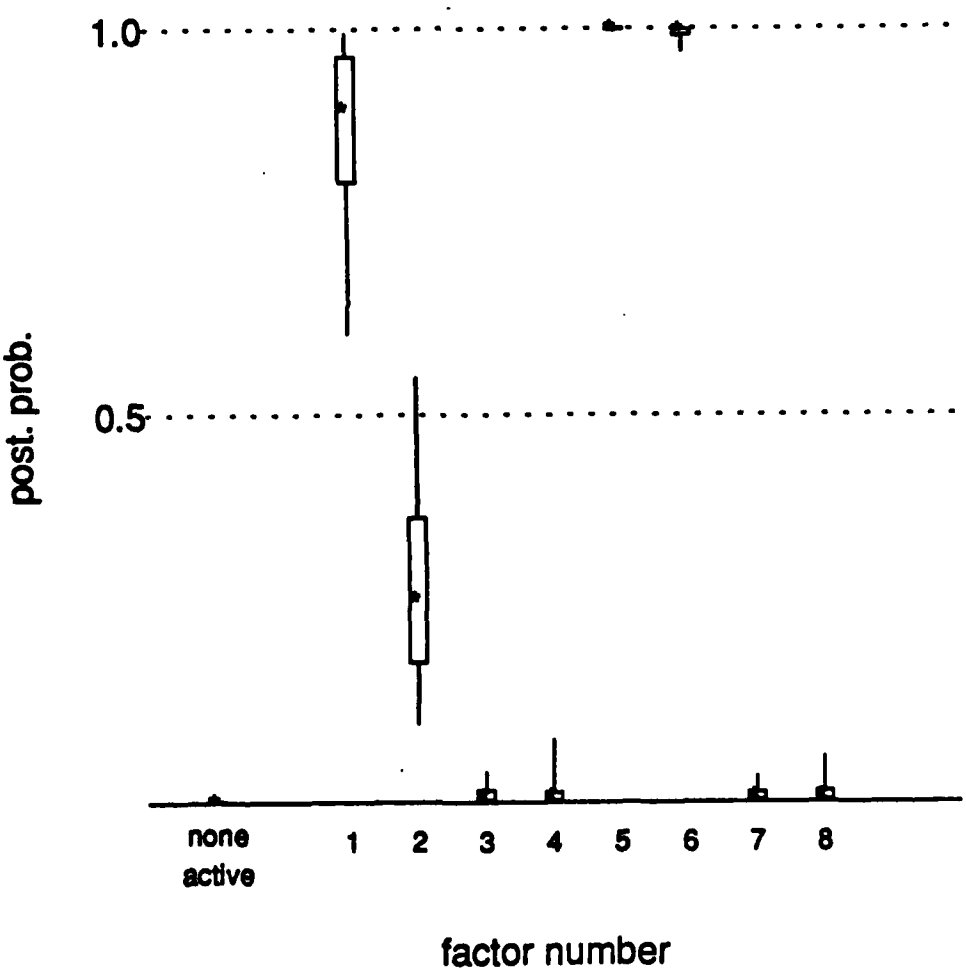


Figure 3.6c Plot of posterior probabilities $\{p_i^*\}$ over 100 simulations, t_3 errors, σ=1, α=0.3, k_1=11, k_2=3.3, five active columns from Example 2.1, factors 1, 2, 5 and 6 possibly active. See Figure 3.2 for explanation of plotting symbols.

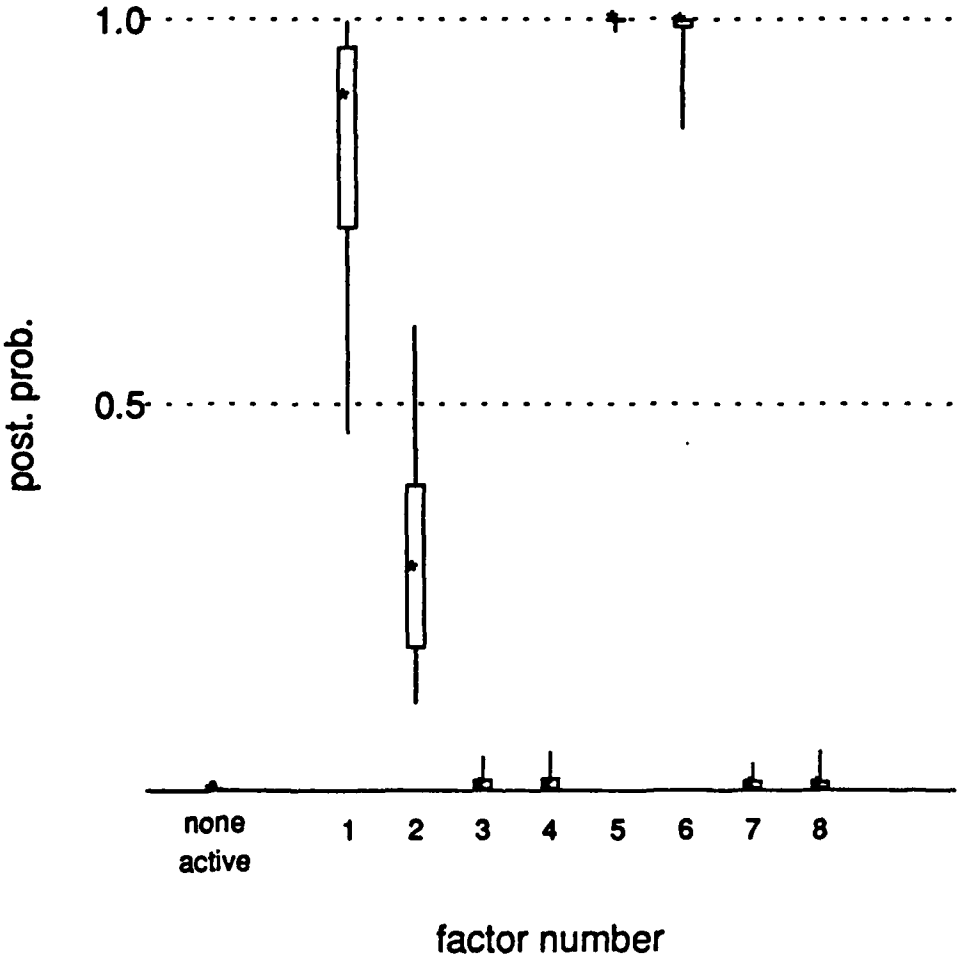
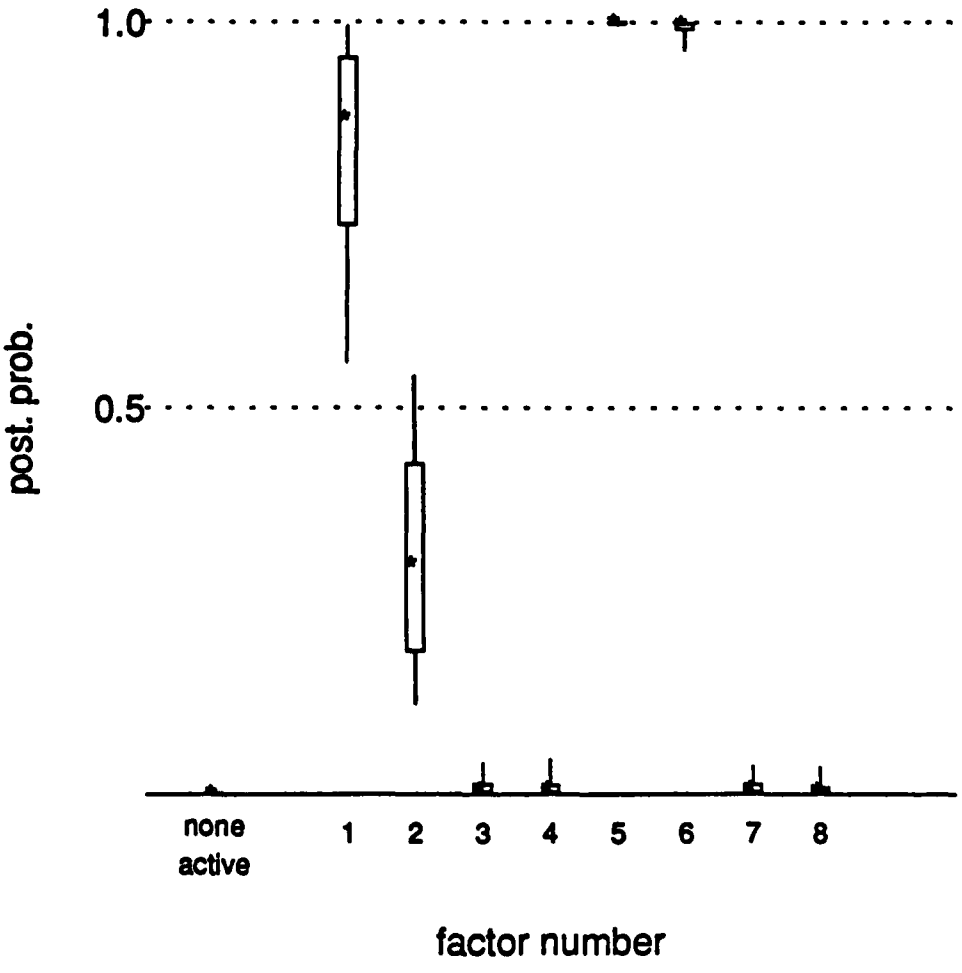


Figure 3.6d Plot of posterior probabilities $\{p_i^*\}$ over 100 simulations, skew normal errors, σ=1, α=0.3, k_1=11, k_2=3.3, five active columns from Example 2.1, factors 1, 2, 5 and 6 possibly active. See Figure 3.2 for explanation of plotting symbols.



how conclusions depend upon these assumptions.

The robustness study gave very satisfying results. The performance of the procedure is not heavily affected by a non-normal error distribution.

CHAPTER 4

BAD OBSERVATIONS IN FACTORIAL EXPERIMENTS

4.1. Introduction

Daniel (1959) has estimated that the frequency of mistaken or bad observations in factorial experiments can reach 10% or higher. Identification of these bad observations is especially difficult for unreplicated experiments, when they may be hidden by biased contrasts wrongly identified as active. The Bayesian methods described in earlier chapters are extended here to accomodate the possibility of bad observations, and compared with the test (John, 1978) for bad values based on the reduction in the residual sum of squares when suspected observations are deleted.

4.2. Fixed Model

John (1978) employs the following model for testing for suspected outliers in fractional factorials. He supposes the general model

$$\mathbf{y} = \mathbf{Z}\tau + \mathbf{e} \tag{4.1}$$

has been fit by least squares, where $\mathbf{Z}$ is the matrix of columns of the design $\mathbf{X}$ corresponding to plausibly active effects. If all columns of $\mathbf{X}$ could plausibly be active then active columns must be identified in some way before continuing, and $\mathbf{Z}$ defined to be the active columns. Define $\mathbf{r}_y$ to be the vector of residuals obtained from the least squares fit of $\mathbf{y}$. If there are m outliers suspected the model is rewritten

$$\mathbf{y} = \mathbf{Z}\tau + \sum_{i=1}^{m} \theta_i \mathbf{d}_i + \mathbf{e} \ , \tag{4.2}$$

where $\mathbf{d}_i$ is a vector with 1 in the row corresponding to the ith bad value, and θ_i is the (unknown) bias in this value. Suppose each of the models

$$\mathbf{d}_i = \mathbf{Z}\tau + \mathbf{e} \quad i = 1,\ldots,m \tag{4.3}$$

is fitted by least squares, and let $\mathbf{R}$ be the $m \times m$ matrix whose (ij)th element is the residual from the least squares fit of $\mathbf{d}_i$ corresponding to the jth suspected outlier of $\mathbf{y}$. Then the portion of the original sum of squares $\mathbf{r}_y{}'\mathbf{r}_y$ due to the m suspected bad values is

$$SS_m = \mathbf{r}_m{}'\mathbf{R}^{-1}\mathbf{r}_m \ , \tag{4.4}$$

where $\mathbf{r}_m$ is the vector of elements of $\mathbf{r}_y$ corresponding to the m supposed bad values. The sum of squares SS_m can be compared to the new resdidual sum of squares $\mathbf{r}_y{}'\mathbf{r}_y - SS_m$ in an F-type test of significance, and for m=1 or 2, John (1978) gives details of the correction to the significance probability for selecting the largest residuals for testing.

For a two-level orthogonal design with

$$\mathbf{Z}'\mathbf{Z} = n\,\mathbf{I}_p \ ,$$

p the number of columns of $\mathbf{Z}$, and one suspected bad value, the sum of squares for outliers is

$$SS_1 = \frac{r_i{}^2}{1 - p/n} \tag{4.5}$$

where r_i is the residual corresponding to the suspected bad value.

John's method is conceptually simple and quite easy to use for one possible bad value. The test statistic is computed easily and the significance probability can also be estimated easily. However, for two suspected bad values, obtaining the maximum SS_2 and its estimated significance probability is more difficult, and a subsequent test is necessary to determine if both or only one of the suspected bad values is actually an outlier. John presents no theory for the possibility of more than two bad values.

There are some drawbacks to the approach. The possibility of more than one bad value is handled only with difficulty or not at all. The method also depends upon a fixed model identified in advance which may be in error due to the presence of bad observations. This is often a minor error as the model can be corrected if and when bad observations are identified.

4.3. A Bayesian Approach

Box and Tiao (1968) detailed a Bayesian approach to the outlier problem in conjunction with the use of the scale-contaminated normal error distribution, which can be written

$$(1-\alpha_2)N(0,\sigma^2)+\alpha_2 N(0,k_2{}^2\sigma^2)\,.$$

With high probability $1-\alpha_2$ an observation is generated by the usual normal model with variance σ^2, and with small probability α_2 the observation has much larger variance $k_2{}^2\sigma^2$. (The subscripts on α_2 and k_2 are to distinguish them from the parameters α and k used previously and henceforth denoted by α_1 and k_1). This outlier model

can be incorporated into the models employed in previous chapters for analyzing factorial experiments. Below are described the details for extending the model for determining active column contrasts, Chapter 2. Extending the model for active factors, Chapter 3, is exactly analogous.

Let α_1 be the prior probability that a particular column is active, and α_2 is the probability that a particular element of y has inflated variance $k_2^2\sigma^2$. The following notation is used in describing the model:

$a_{(c)}$ = event that a particular set of c columns is active.

$a_{(r)}$ = event that a particular set of r observations are "outliers", i.e., have inflated variance.

$a_{(r,c)} = a_{(r)} \cap a_{(c)}$.

$\mathbf{X}_{(c)}$ = columns of $\mathbf{X}$ corresponding to $a_{(c)}$.

$\tau_{(c)}$ = vector of true contrasts corresponding to $a_{(c)}$.

$\mathbf{X}_{(r,c)}$ = rows and columns of $\mathbf{X}$ corresponding to active columns and bad values of $a_{(r,c)}$.

$\mathbf{y}_{(r)}$ = rows of y corresponding to bad values of $a_{(r,c)}$.

$\phi_2 = 1 - 1/k_2^2$.

Then, the sampling distribution of y, given $a_{(r,c)}$, is

$$p(y \mid \tau,\sigma^2,a_{(r,c)}) = (2\pi)^{-n/2}\sigma^{-n}k_2^{-r} \times \tag{4.6}$$

$$\exp\left\{\frac{-1}{2\sigma^2}\left[(y - X_{(c)}\tau_{(c)})'(y - X_{(c)}\tau_{(c)}) - \right.\right.$$

$$\left.\left.\phi_2(y_{(r)} - X_{(r,c)}\tau_{(c)})'(y_{(r)} - X_{(r,c)}\tau_{(c)})\right]\right\}.$$

The prior distribution of $\tau_{(c)}$ and σ are again given by (2.2). The posterior distribution of $\tau_{(c)}$ given $a_{(r,c)}$ is

$$p(\tau_{(c)} \mid a_{(r,c)},y) \;\alpha\; \int_0^\infty \sigma^{-n-c-1}\gamma^{-c}k_2^{-r} \times \tag{4.7}$$

$$\exp\left\{\frac{-1}{2\sigma^2}\left[S_{(r,c)}(\tau_{(c)}) + \tau_{(c)}'\Gamma_c\tau_{(c)}\right]\right\}d\sigma$$

$$= \gamma^{-c}k_2^{-r}\left[S_{(r,c)}(\tau_{(c)}) + \tau_{(c)}'\Gamma_c\tau_{(c)}\right]^{-(n+c)/2},$$

where

$$S_{(r,c)}(\tau_{(c)}) = (y - X_{(c)}\tau_{(c)})'(y - X_{(c)}\tau_{(c)}) \tag{4.8}$$

$$-\phi_2(y_{(r)} - X_{(r,c)}\tau_{(c)})'(y_{(r)} - X_{(r,c)}\tau_{(c)}).$$

Now let

$$G_{(r,c)} = \Gamma_c + X_{(c)}'X_{(c)} - \phi_2 X_{(r,c)}'X_{(r,c)} \tag{4.9}$$

and

$$\hat{\tau}_{(r,c)} = G^{-1}_{(r,c)}(X_{(c)}'y - \phi_2 X_{(r,c)}'y_{(r)}) . \tag{4.10}$$

Then it follows that

$$S_{(r,c)}(\tau_{(c)}) + \tau_{(c)}'\Gamma_c\tau_{(c)} = \tag{4.11}$$

$$(\tau_{(c)} - \hat{\tau}_{(r,c)})'G_{(r,c)}(\tau_{(c)} - \hat{\tau}_{(r,c)}) + S_{(r,c)}(\hat{\tau}_{(r,c)}) + \hat{\tau}_{(r,c)}'\Gamma_c\hat{\tau}_{(r,c)} .$$

Thus the posterior density of $\tau_{(c)}$ given $a_{(r,c)}$ must be

$$p(\tau_{(c)}|a_{(r,c)},y) = \frac{\Gamma((n+c-1)/2)}{\Gamma((n-1)/2)} \frac{|G_{(r,c)}|^{1/2}}{(\pi(n-1)s_{(r,c)}^2)^{1/2}} \times \tag{4.12}$$

$$\left\{1 + \frac{(\tau_{(c)} - \hat{\tau}_{(r,c)})'G_{(r,c)}(\tau_{(c)} - \hat{\tau}_{(r,c)})}{(n-1)s_{(r,c)}^2}\right\}^{-(n+c)/2} ,$$

which is a $(c+1)$-dimensional multivariate t density with $n-1$ degrees of freedom, mean vector $\hat{\tau}_{(r,c)}$, and dispersion matrix $s_{(r,c)}^2 G_{(r,c)}$, where

$$s_{(r,c)}^2 = \frac{S_{(r,c)}(\hat{\tau}_{(r,c)}) + \hat{\tau}_{(r,c)}'\Gamma_c\hat{\tau}_{(r,c)}}{n-1} . \tag{4.13}$$

Then, following equations (2.3)-(2.5), the posterior probability of the event $a_{(r,c)}$ can be written

$$p(a_{(r,c)}|y) \propto \left[\frac{\alpha_1}{1-\alpha_1}\right]^c \left[\frac{\alpha_2}{1-\alpha_2}\right]^r \gamma^{-c} k_2^{-r} \times \frac{|G_{(0,0)}|^{1/2}}{|G_{(r,c)}|^{1/2}} \left[\frac{s_{(r,c)}^2}{s_{(0,0)}^2}\right]^{-(n-1)/2}, \tag{4.14}$$

which depends mainly on the prior probability of the event and the sum of squared residuals when the supposed bad observations are downweighted by the factor k_2. Then the posterior probability that column i is active is

$$p_i = \sum_{(r,c):\, i \text{ active}} p(a_{(r,c)}|y) \tag{4.15}$$

and the probability that observation y_j is bad is

$$q_j = \sum_{(r,c):\, j \text{ bad}} p(a_{(r,c)}|y) \ . \tag{4.16}$$

4.3.1. Some Computational Aspects

To obtain the probabilities $\{p_i\}$, $i=1,...,n-1$, and $\{q_j\}$, $j=1,...,n$, the probabilities $p(a_{(r,c)}|y)$ must be computed for all 2^{2n-1} combinations of possible active columns and bad observations. For $n=16$ there are over two billion such combinations. However, the grand majority of these will have negligible posterior probability. Then, for example, attention may be restricted to events $a_{(r,c)}$ with r and c less than some reasonable upper bounds. The number of possible bad observations especially could be reasonably assumed to be less than two or three in most cases for a 16-run design. Once this assumption is made, most events of interest will have $r < c$, so that

the following derivations will also reduce the computing time somewhat.

The following identity follows from the formula for the inverse of a sum of matrices:

$$\mathbf{G}_{(r,c)}^{-1} = \mathbf{G}_{(0,c)}^{-1} + \phi_2 \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}' \left[\mathbf{I} - \phi_2 \mathbf{X}_{(r,c)} \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}' \right]^{-1} \mathbf{X}_{(r,c)} \mathbf{G}_{(0,c)}^{-1} .$$

It then follows that

$$\hat{\tau}_{(r,c)} = \hat{\tau}_{(0,c)} - \phi_2 \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}' \left[\mathbf{I} - \phi_2 \mathbf{X}_{(r,c)} \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}' \right]^{-1} (\mathbf{y}_{(r)} - \mathbf{X}_{(r,c)} \hat{\tau}_{(r,c)}) . \tag{4.17}$$

If Δ is defined by

$$\Delta = \left[\mathbf{I} - \phi_2 \mathbf{X}_{(r,c)} \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}' \right]^{-1} (\mathbf{y}_{(r)} - \mathbf{X}_{(r,c)} \hat{\tau}_{(0,c)}) , \tag{4.18}$$

then

$$\hat{\tau}_{(r,c)} = \hat{\tau}_{(0,c)} - \phi_2 \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}' \Delta . \tag{4.19}$$

Note that Δ is the solution of a $r \times r$ linear system involving the residuals corresponding to suspected bad observations, obtained from the calculations assuming no bad values, whereas $\hat{\tau}_{(r,c)}$ previously was defined as the solution of a $(c+1)\times(c+1)$ system. The matrix inverse $\mathbf{G}_{(0,c)}^{-1}$ in (4.18) can be obtained easily because the matrix $\mathbf{G}_{(0,c)}$ is diagonal. The $(c+1)\times(c+1)$ determinant $|\mathbf{G}_{(r,c)}|$ in the expression (4.13) for $p(a_{(r,c)}|\mathbf{y})$ must still be computed. This determinant can be related to the deter-

minant

$$|\mathbf{I} - \phi_2 \mathbf{X}_{(r,c)} \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}'| ,$$

which is the determinant of the $r \times r$ system from which Δ is computed.

Note that

$$\mathbf{G}_{(r,c)} = \mathbf{G}_{(0,c)} \left[\mathbf{I} - \phi_2 \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}' \mathbf{X}_{(r,c)} \right]$$

so that

$$|\mathbf{G}_{(r,c)}| = |\mathbf{G}_{(0,c)}| \; |\mathbf{I} - \phi_2 \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}' \mathbf{X}_{(r,c)}| .$$

Rao (1973) shows for A, D nonsingular, not necessarily of the same dimension,

$$\det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = |\mathbf{A}| \; |\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}| = |\mathbf{D}| \; |\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}| .$$

Therefore let

$$\mathbf{A} = -\mathbf{G}_{(0,c)}/\phi_2 , \quad \mathbf{B} = \mathbf{X}_{(r,c)}'$$
$$\mathbf{C} = -\mathbf{X}_{(r,c)} , \quad \mathbf{D} = \mathbf{I}_r .$$

This implies that

$$|-\mathbf{G}_{(0,c)}/\phi_2| \; |\mathbf{I}_r - \phi_2 \mathbf{X}_{(r,c)} \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}'|$$
$$= |-\mathbf{G}_{(0,c)}/\phi_2 + \mathbf{X}_{(r,c)}' \mathbf{X}_{(r,c)}| ,$$

which implies

$$|\mathbf{I}_r - \phi_2 \mathbf{X}_{(r,c)} \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}'| = |\mathbf{I}_{c+1} - \phi_2 \mathbf{G}_{(0,c)}^{-1} \mathbf{X}_{(r,c)}' \mathbf{X}_{(r,c)}|$$

so that

$$|\mathbf{G}_{(r,c)}| = |\mathbf{G}_{(0,c)}|\ |\mathbf{I}_r - \phi_2 \mathbf{X}_{(r,c)} \mathbf{G}^{-1}_{(0,c)} \mathbf{X}_{(r,c)}'|\ . \tag{4.20}$$

The implication of these identities is that for each combination of c active contrasts assessed, the probabilities $p_{(r,c)}$ can be obtained from $p_{(0,c)}$ over all values of r by solving a $r \times r$ linear system rather than a $c+1 \times c+1$ system.

4.4. Example

The following example is taken from Box and Draper (1986).

Example 4.1 Four factors were studied in a full 2^4 factorial experiment. The design array and observations are presented in Table 4.1. Applying the Bayesian analysis of Chapter 2 to identify the active contrasts of the experiment, with α=0.2, k=10, the posterior probabilities are plotted in Figure 4.1. Main effects of factors 2 and 3 stand out from the other effects, but evidence for their activity is dubious because of the relatively low probabilities obtained. A plot of the residuals, Figure 4.2, taken after including terms for main effects 2 and 3 as well as the mean, reveals that the residual corresponding to observation y_{13}=59.15 is much larger in absolute value than the others. A normal plot of the contrasts, Figure 4.3, reveals a gap in the cluster of contrasts near zero, another sign (Daniel (1959); see Chapter 1) that there is a possibly bad observation.

Applying the Bayesian analysis now allowing for the possibility of bad values, with α_2=.05 and k_2=5, the posterior probabilities $\{q_j\}$ of observations being bad and $\{p_i\}$ of contrasts being active were computed and are plotted in Figure 4.4. The values of α_2 and k_2 chosen for illustration were suggested as moderate values by Chen and Box (1979). (The computations were carried out assuming there were six or fewer active contrasts and two or fewer bad values, an event of prior probability .94). The value of q_{13} is very close to one, suggesting strongly that observation y_{13} is bad. The affect on the probabilities $\{p_i\}$ of the automatic downweighting of y_{13} achieved by the Bayesian analysis is to make the posterior probabilities for main effects 2 and 3

Table 4.1 Design array and observations for Example 4.1, a full 2^4 factorial experiment from Box and Draper (1986).

	factors				
run	1	2	3	4	y
1	-	-	-	-	47.46
2	+	-	-	-	49.62
3	-	+	-	-	43.13
4	+	+	-	-	46.31
5	-	-	+	-	51.47
6	+	-	+	-	48.49
7	-	+	+	-	49.34
8	+	+	+	-	46.10
9	-	-	-	+	46.76
10	+	-	-	+	48.56
11	-	+	-	+	44.83
12	+	+	-	+	44.45
13	-	-	+	+	59.15
14	+	-	+	+	51.33
15	-	+	+	+	47.02
16	+	+	+	+	47.90

column(effect)	observed contrast	column(effect)	observed contrast
0(mean)	48.25	8(4)	1.01
1(1)	-0.80	9(14)	-0.58
2(2)	-4.22	10(24)	-1.18
3(12)	0.91	11(124)	0.72
4(3)	3.71	12(34)	-1.49
5(13)	-2.49	13(134)	0.40
6(23)	-0.80	14(234)	-1.58
7(123)	1.20	15(1234)	1.52

Figure 4.1 Plot of posterior probabilities $\{p_i\}$ for Example 4.1, α=0.2, k=10, assuming no bad values.

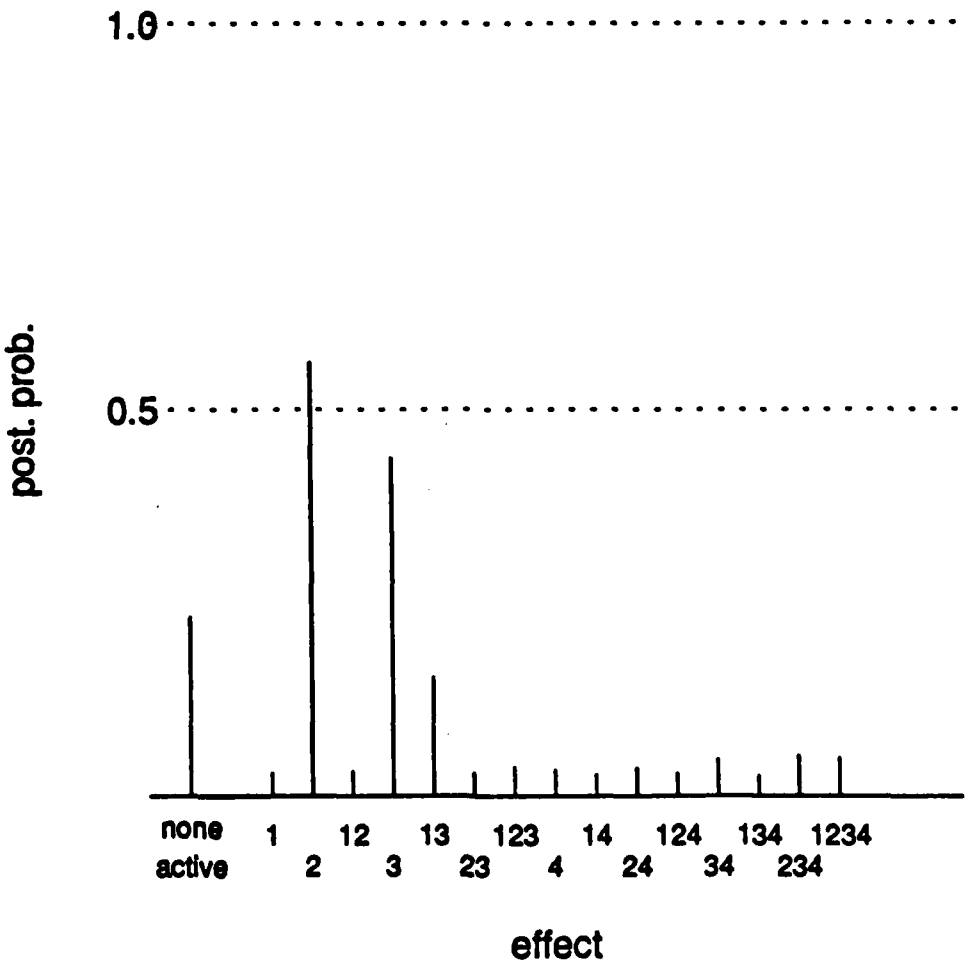


Figure 4.2 Plot of residuals versus run order after fitting model with 2, 3 main effects, Example 4.1.

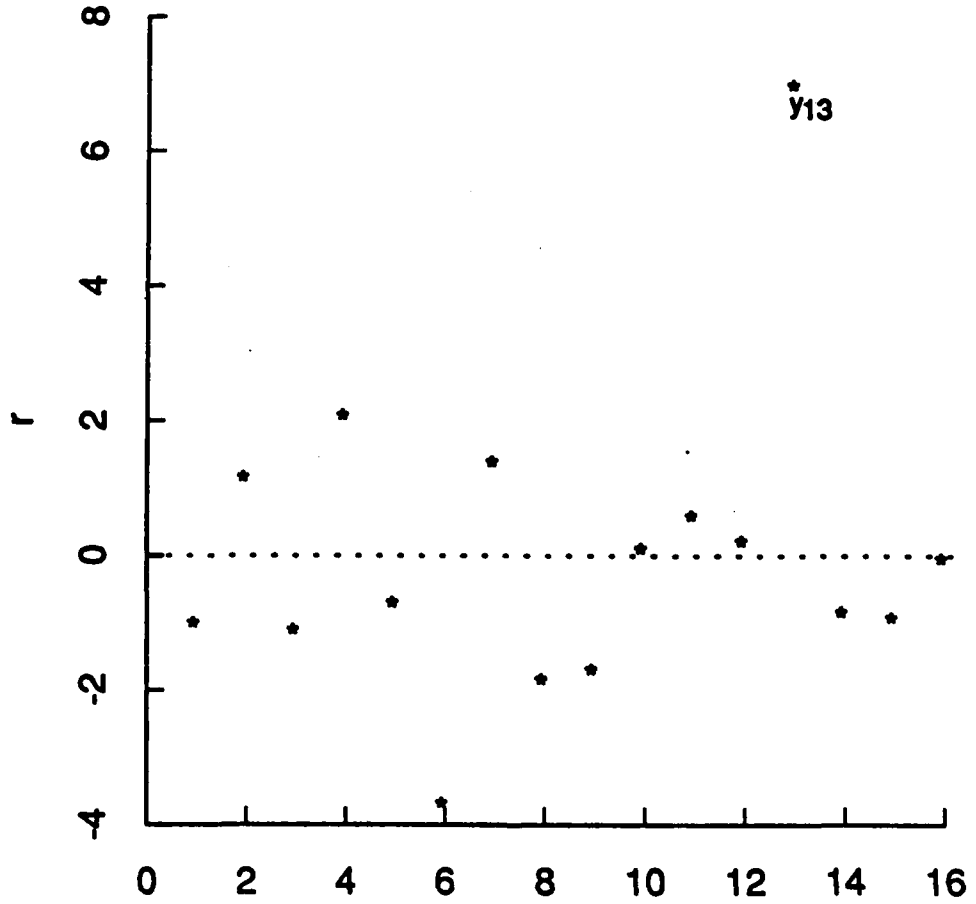


Figure 4.3 Full normal plot of observed contrasts, Example 4.1.

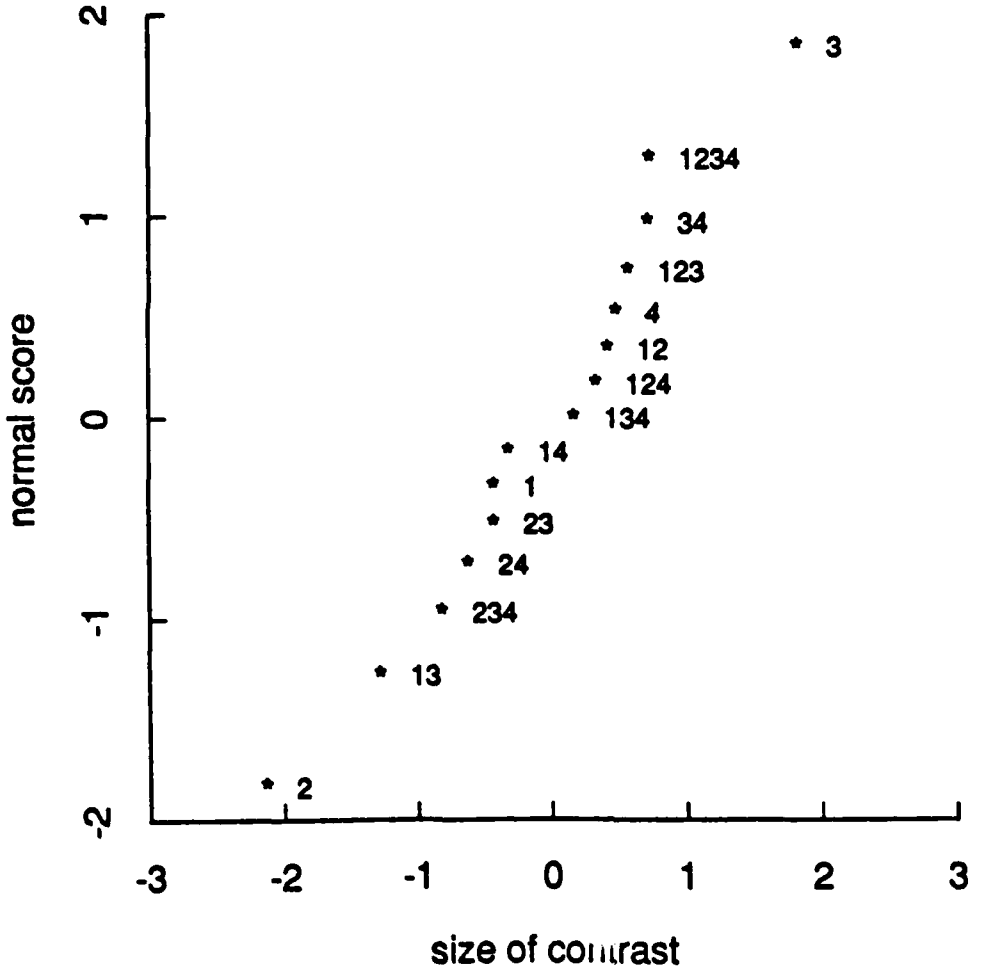


Figure 4.4a Plot of posterior probabilities $\{q_j\}$ that observations are bad, Example 4.1, α_1=0.2, k_1=10, α_2=0.05, k_2=5.

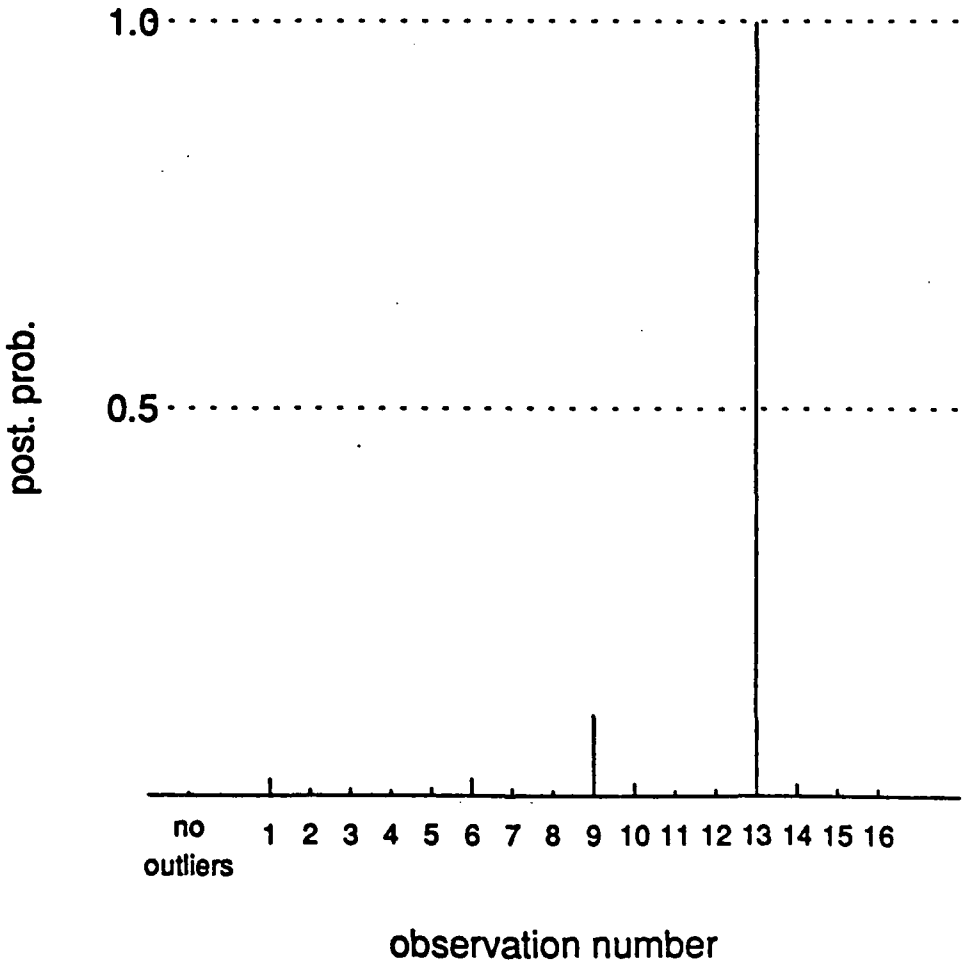
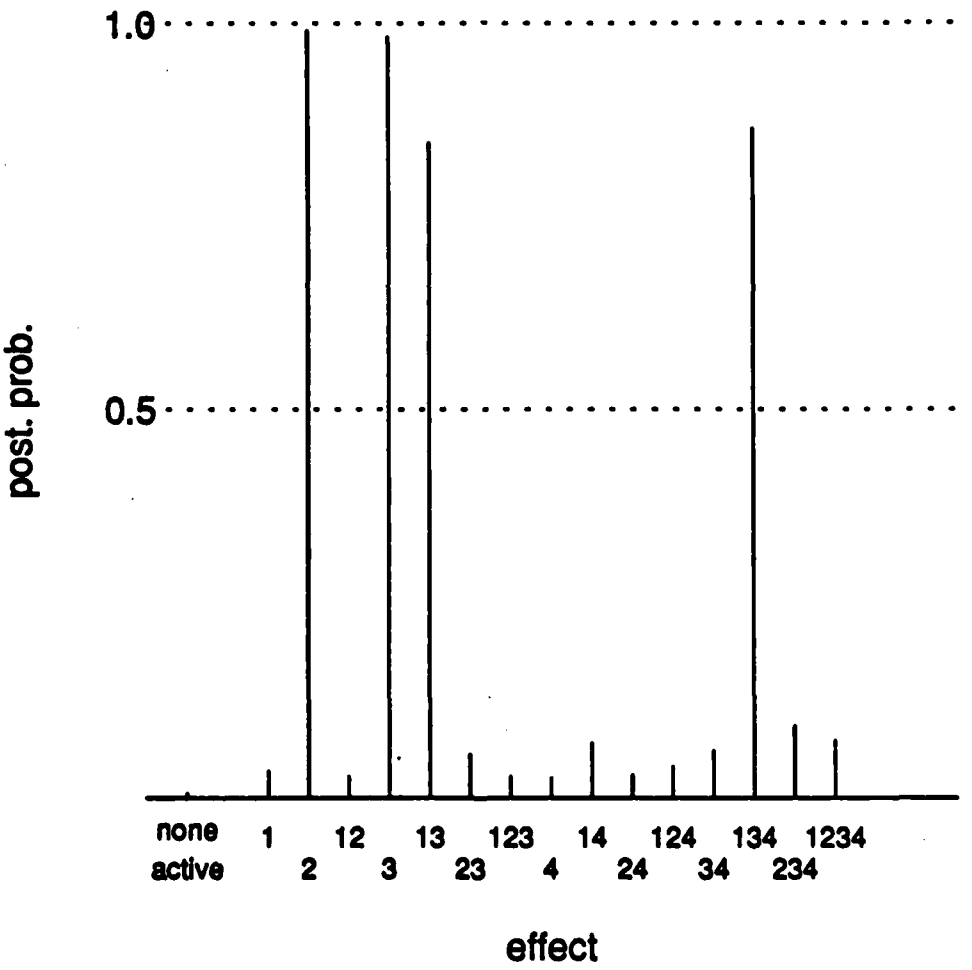


Figure 4.4b Plot of posterior probabilities $\{p_i\}$ that contrasts are active, Example 4.1, allowing for possible bad observations, α_1=0.2, k_1=10, α_2=0.05, k_2=5.



much closer to one. The 13 and 134 interactions, previously hidden, also receive much higher posterior probabilities.

Applying John's method for testing for the significance of the possible outlier y_{13}, an F-statistic of 33.04 on 1 and 12 degrees of freedom is obtained. The estimated significance probability, correcting for selection of the largest residual, is .0015. Thus y_{13} is judged to be an outlier, and assuming subsequent investigation did not reveal the correct value for this observation, it is replaced by its least squares missing value estimate of $\dot{y}_{13}$=51.56. The posterior probabilities $\{p_i\}$ were recomputed based on the revised data pretending $\dot{y}_{13}$ was actually observed, and are plotted in Figure 4.5. The values are very close to those obtained from the complete Bayesian analysis in Figure 4.4.

It has been demonstrated that for this example testing for bad values by examination of residuals after a model has been identified leads to the same conclusions as the simultaneous identification of active contrasts and bad observations via the Bayesian analysis described earlier. However, the observation identified as bad in this example was so far removed from the pattern of the rest of the data that any procedure which failed to flag it would be a poor one indeed. Thus the procedures compared here meet this minimum standard. Suppose now the observation y_{13} is replaced by a value somewhere between the original value of 59.15 and 51.56, namely let y_{13}=55.15. The data shall be reanalyzed by both methods and the results compared.

The Bayesian analysis applied to the new data, not allowing for bad observations, leads to the posterior probabilities $\{p_i\}$ plotted in Figure 4.6. Main effects of

Figure 4.5 Plot of posterior probabilities $\{p_i\}$ that contrasts are active, assuming no bad observations, Example 4.1 with y_{13} replaced by its least squares missing value estimate.

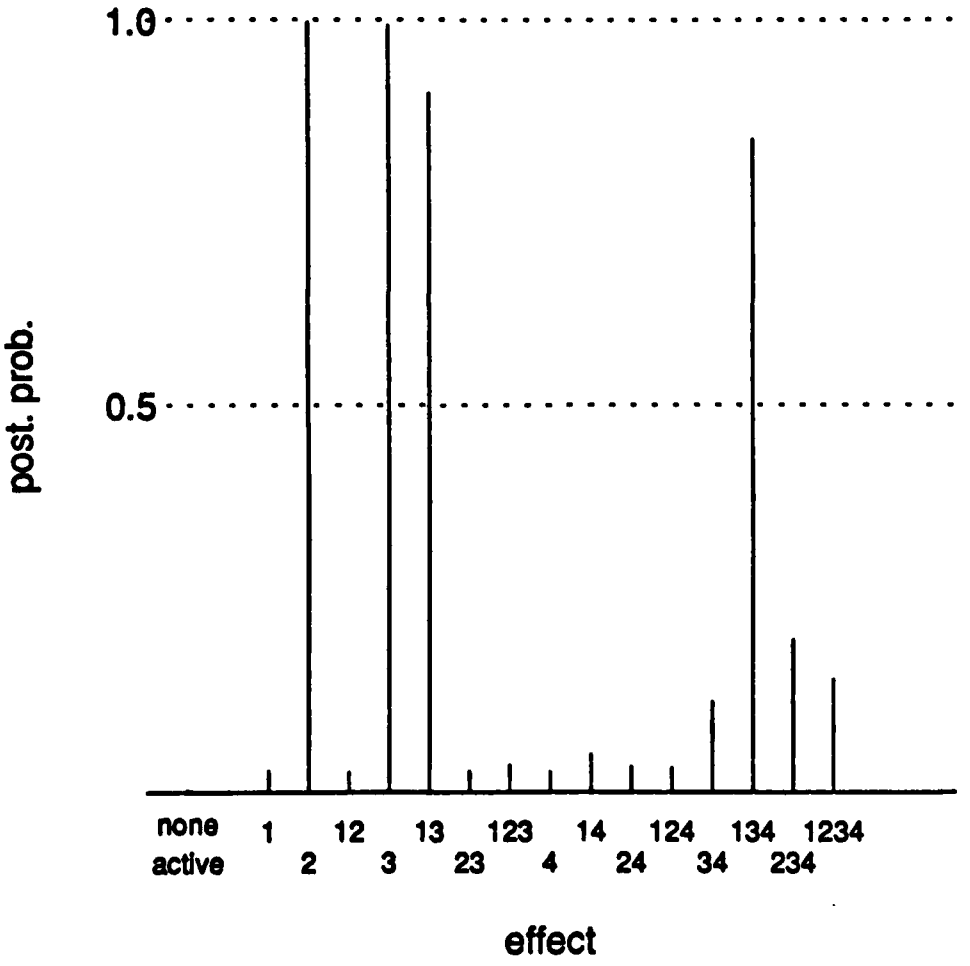
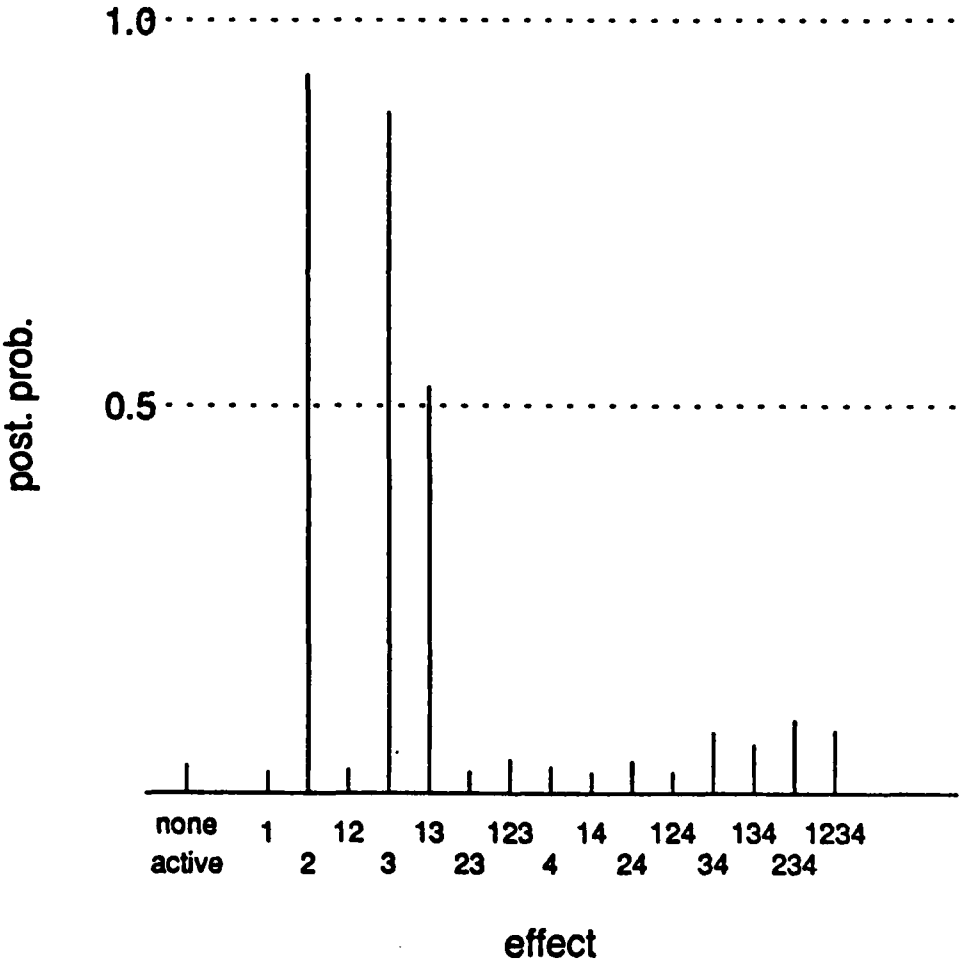


Figure 4.6 Plot of posterior probabilities $\{p_i\}$ that contrasts are active, assuming no bad observations, Example 4.1 with y_{13}=55.15 (midway between the original value and the missing value estimate).



variables 2 and 3 are apparently active, and there is some evidence for the activity of the 13 interaction. A normal plot of the contrasts, Figure 4.7, agrees with this assessment, but the gap which appeared among the apparently inert contrasts for the original data has now largely disappeared. A plot of the residuals after fitting a model with the three possibly active effects indicated above, Figure 4.8, again reveals the residual corresponding to observation y_{13} is the largest in absolute magnitude. Application of the F-test for rejection of observation y_{13} as bad yields an F-ratio of 8.35 on 1 and 12 degrees of freedom and an estimated significance probability of .24. (Applying the test to the residuals after fitting a model with main effects 2 and 3 only, gave an estimated significance probability of .16). Thus the observation y_{13} would not be rejected as bad and the conclusions about active effects stated above would hold barring any further developments from diagnostic checking.

Applying the Bayesian analysis to the new data allowing for possible bad observations, with α_1=0.2, α_2=0.05, k_1=10 and k_2=5, the posterior probabilities $\{p_i\}$ and $\{q_j\}$ are plotted in Figure 4.9. The plot of the column posterior probabilities reveals that when the possibility of bad observations is taken into account, there is evidence for the activity of the previously hidden 134 interaction, as well as stronger evidence for the other three effects identified previously, Figure 4.6. The plot of the posterior probabilities of observations being bad shows there is substantial evidence that the observation y_{13} is faulty. The F-test for bad values failed to identify it as bad because the fixed model was misspecified due to the presence of the bad observation. If the model had been specified to include the 134 interaction and the F-test for bad values

Figure 4.7 Full normal plot of observed contrasts, Example 4.1 with y_{13}=55.15 (midway between the original value and the missing value estimate).

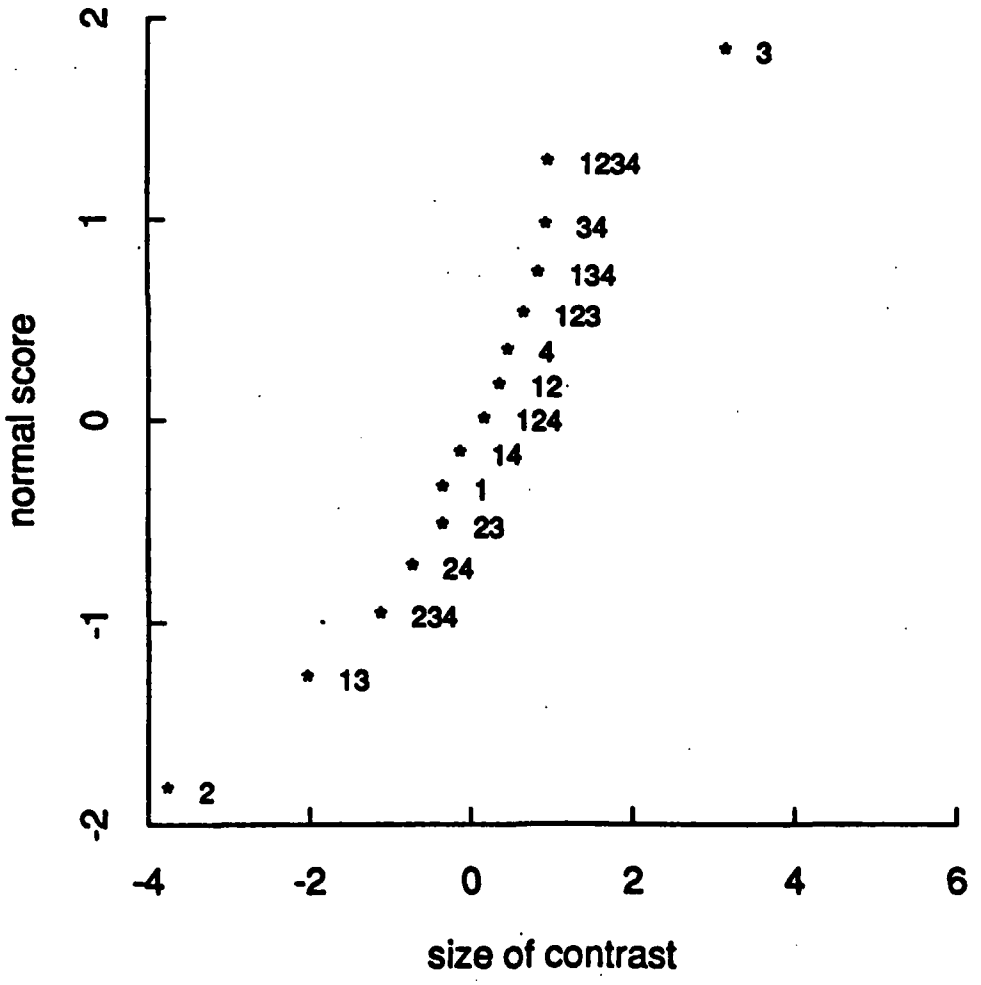


Figure 4.8 Plot of residuals versus run order after fitting model with 2, 3 and 13 effects, Example 4.1 with y_{13}=55.15 (midway between the original value and the missing value estimate).

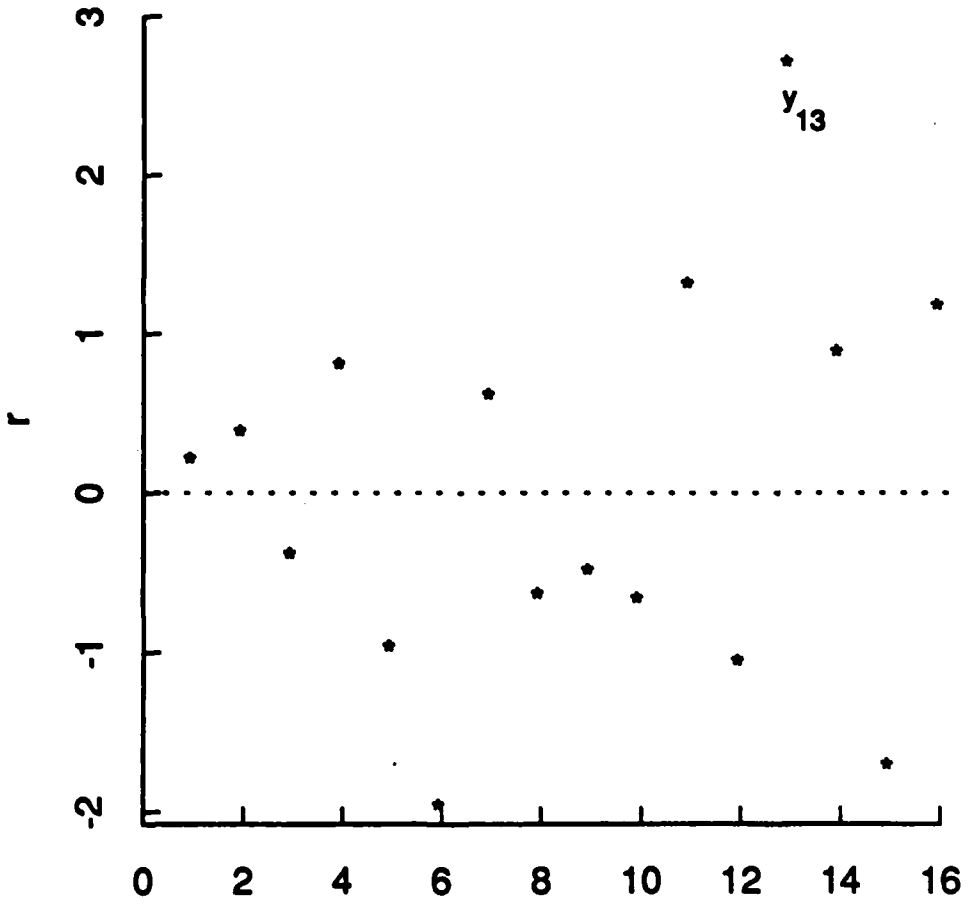


Figure 4.9a Plot of posterior probabilities $\{p_i\}$ that contrasts are active, allowing for possible bad observations, α_1=0.2, k_1=10, α_2=0.05, k_2=5, Example 4.1 with y_{13}=55.15 (midway between the original value and the missing value estimate).

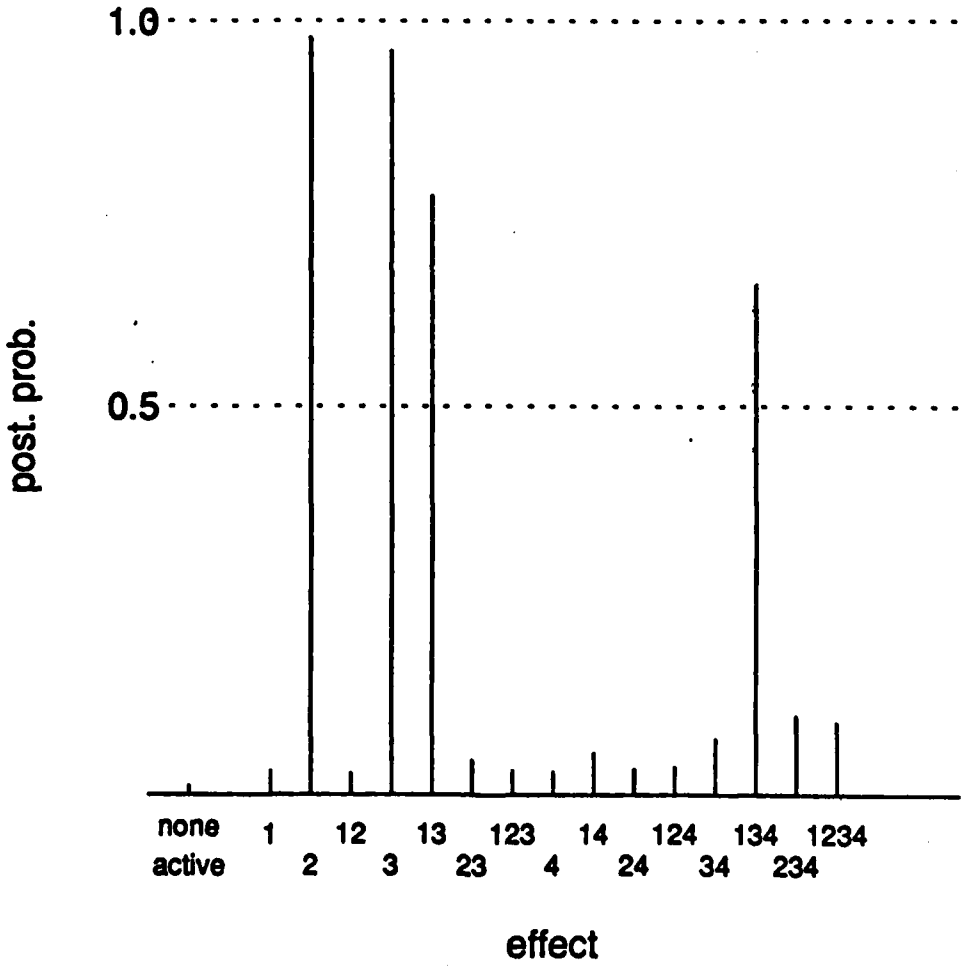
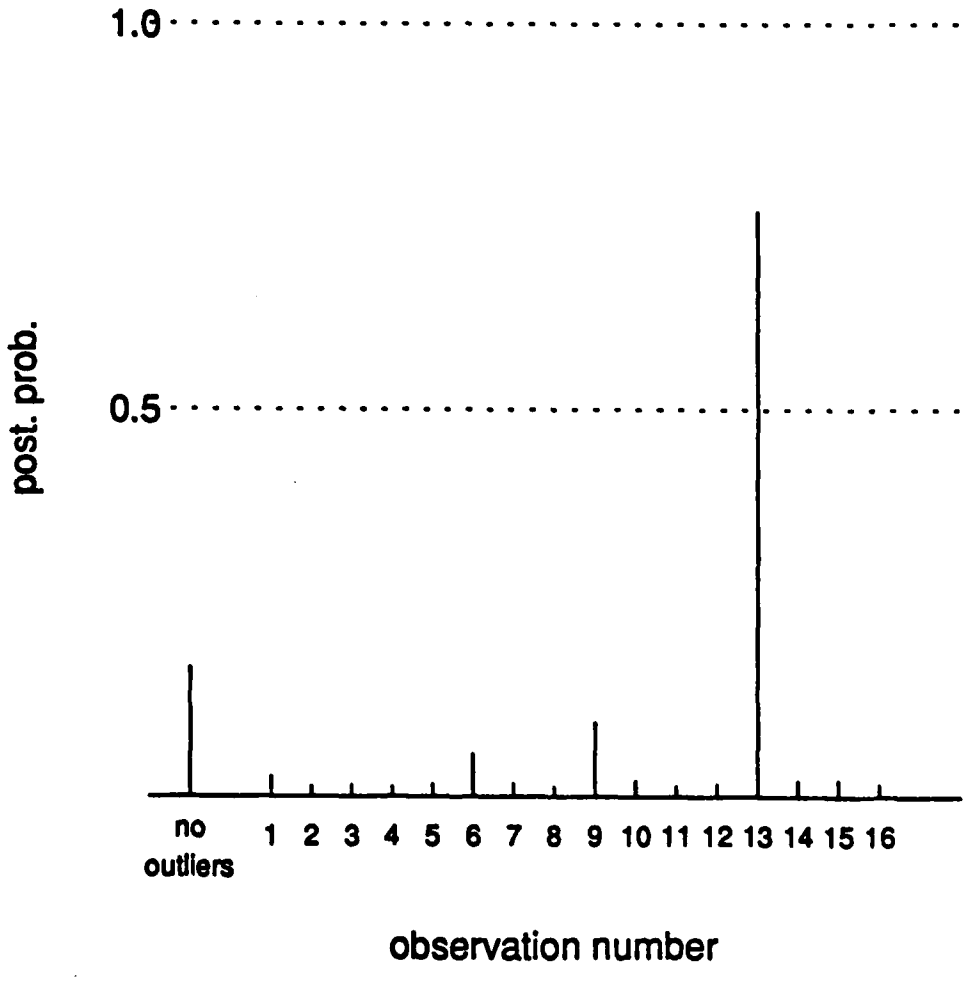


Figure 4.9b Plot of posterior probabilities $\{q_j\}$ that observations are bad, α_1=0.2, k_1=10, α_2=0.05, k_2=5, Example 4.1 with y_{13}=55.15 (midway between the original value and the missing value estimate).



had been carried out on those residuals, an F-ratio of 29.83 on 1 and 10 degrees of freedom would have been obtained with an estimated significance probability of .004.

It is not claimed that the 134 interaction would not have been discovered eventually without the Bayesian analysis, for example by plotting residuals against the various factor levels. It has been shown, however, that straightforward application of existing methodology could have led to incomplete conclusions. Also, it is not suggested that the Bayesian analysis led to the "correct" answer, but it did uncover a plausible explanation of the data, i.e., observation y_{13} might be wrong and the 134 interaction might be active.

The two methods compared here are not that different mathematically in that both assess the possibility of bad observations according to the reduction in the residual sum of squares when observations are deleted, or downweighted as in the Bayesian analysis. The difference comes from the fact that the Bayesian model, in complete generality, assesses all possible combinations of active effects or contrasts and bad observations, whereas the F-test generally does not. The test for bad values could theoretically be applied to all possible models and combinations of bad observations, but this leads to an exceedingly complex repeated-testing problem, whereas the proper weighting of all combinations comes automatically in the Bayesian analysis.

The premium paid for the generality of the Bayesian analysis is a sharp increase in computing requirements. Reasonable assumptions about the number of active contrasts and bad observations helps to reduce these requirements. For example, for the analyses carried out in the previous example, it was assumed there were six or fewer

active contrasts and two or fewer bad observations, eliminating 99.94% of the number of combinations to be considered while losing events of prior probability less than .07. Yet, four hours of computing time were required to compute the posterior probabilities $\{p_i\}$ and $\{q_j\}$.

4.5. Approximating the Posterior Probabilities

An approximation to the full Bayesian analysis is motivated by the (model identification)-(fitting)- (diagnostic checking) iteration discussed, for example, by Box and Jenkins (1976), p. 18. At the first iteration the model identification step would entail identifying active contrasts or factors according to the methods described in earlier chapters, assuming that errors were independent and normally distributed with constant variance. Then, residuals are obtained after fitting the model in the usual way. These are examined for possible departures from model assumptions, and the model respecified if any appear.

At the first iteration then, it is assumed all errors are independent from the $N(0,\sigma^2)$ distribution. The posterior probabilities $\{p_i\}$ that contrasts are active are computed, not allowing for bad values. All contrasts receiving posterior probability greater than some value P are identified as active, P to be chosen possibly after examining $\{p_i\}$. The probabilities $\{q_j\}$ that observations are bad can be computed, conditional on the model fixed at the previous step. Those observations with posterior probabilities greater than Q, Q to be chosen, are assumed to be bad, i.e., they have variance $k_2^2\sigma^2$. If there are no bad observations, the iteration stops. If there are observations identified as bad, the model is respecified by computing the probabilities $\{p_i\}$ conditional on bad observations having larger variance. If the contrasts identified as active at this step are the same as a previous iteration, the iteration stops. If a new set of contrasts is identified as active, the iteration continues with the computation of $\{q_j\}$, etc.

There is no guarantee that this procedure will converge to a state close to the true probabilities $\{p_i\}$ and $\{q_j\}$. It may oscillate; for example, the assumption of no bad values may lead to one set of active contrasts, which leads to the identification of a bad observation, which implies a different set of active contrasts, which in turn implies no bad observations, completing the circle. (This type of oscillation has not been observed in any examples up to this time). Also, the procedure can and often does converge to different states depending on the choice of P and Q. However, it does perform well in discovering various possible explanations of the data, especially when the approximation is repeated for different values of P and Q. In the end, too, competing hypotheses can be compared according to their posterior probability ratio, which can be computed exactly. The procedure is illustrated for the original and revised data of Example 4.1.

Starting with the original data, the first step is to compute the probabilities $\{p_i\}$ assuming no bad values, which was done previously. These are plotted in Figure 4.1. Choosing P=0.4, main effects 2 and 3 are tentatively identified as active. The probabilities $\{q_j\}$ were computed conditional on the identified model, and are plotted in Figure 4.10. Observation y_{13} has posterior probability close to one and any reasonable choice of Q would lead to this observation being identified as bad. The probabilities $\{p_i\}$ were recomputed based on that identification, and are plotted in Figure 4.11. Main effects 2 and 3 now have much higher probability, and the 13 and 134 interactions could now also be identified as active. The $\{q_j\}$ based on these four effects being active are almost indistinguishable from the previous iteration and thus

Figure 4.10 Plot of posterior probabilities $\{q_j\}$ that observations are bad, after one step of iterative approximation, Example 4.1, α_1=0.2, k_1=10, α_2=0.05, k_2=5.

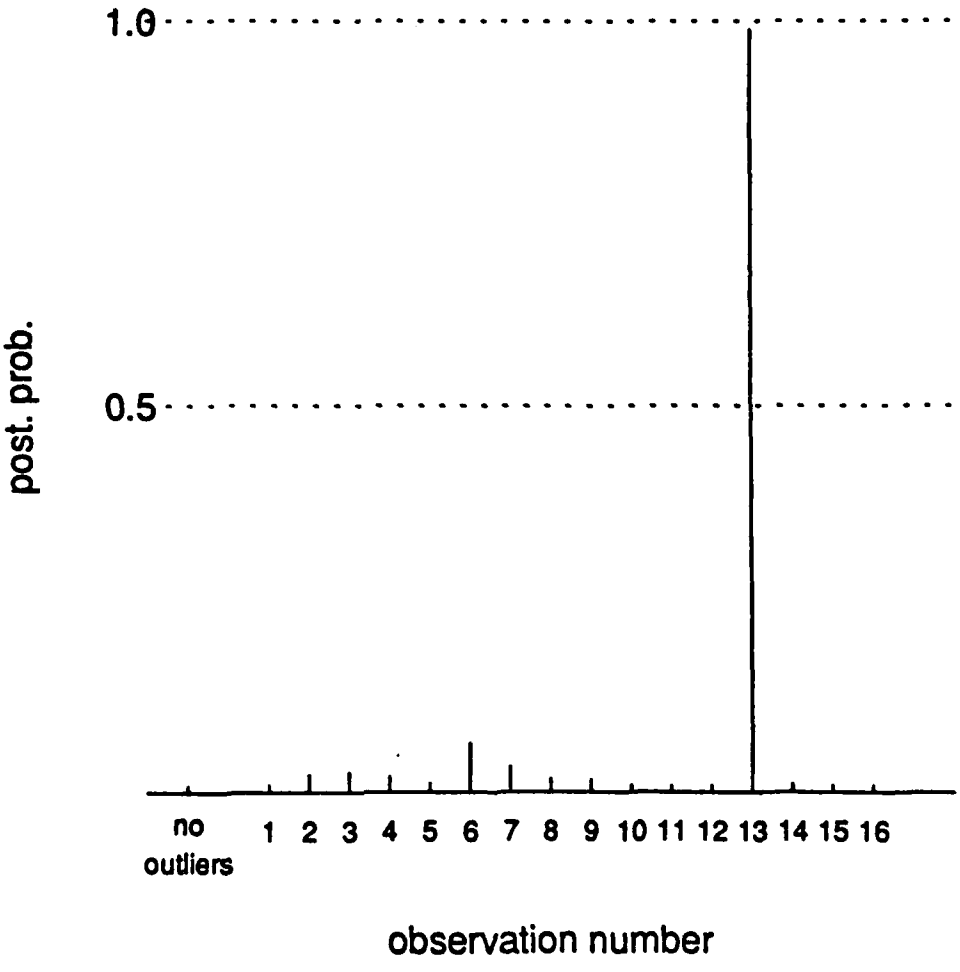
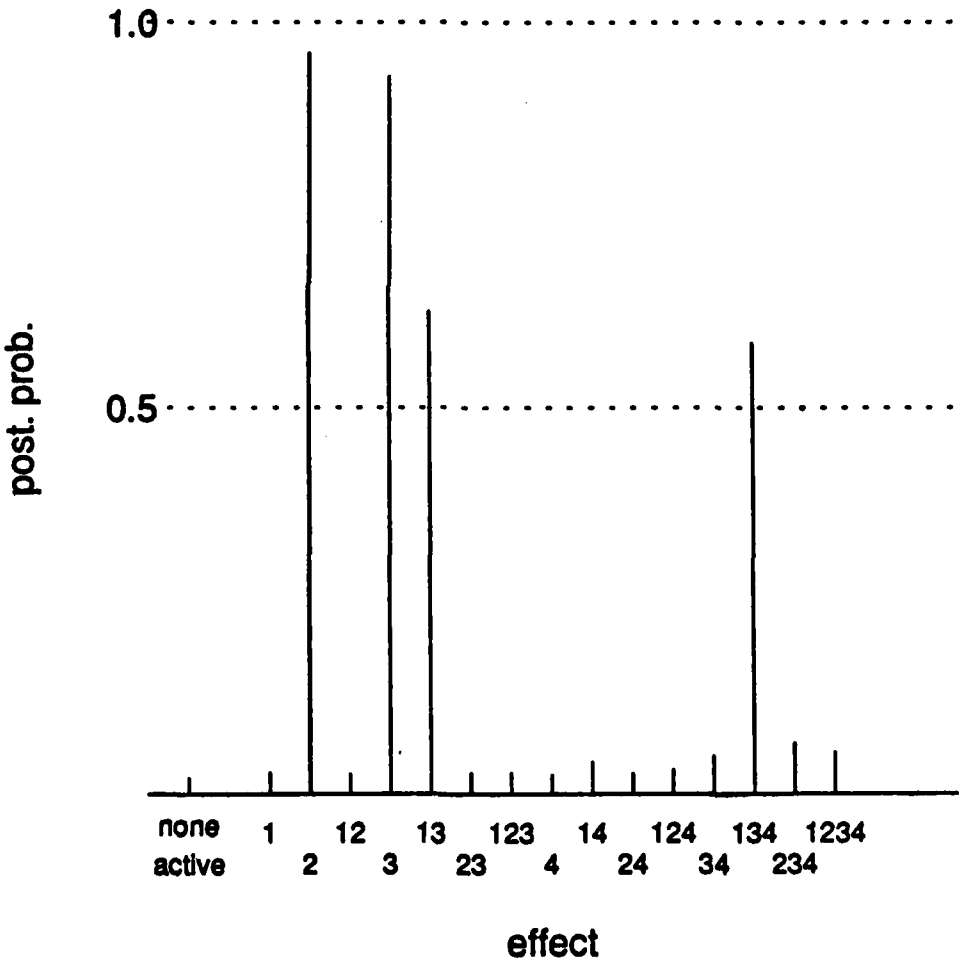


Figure 4.11 Plot of posterior probabilities $\{p_i\}$ that contrasts are active, after two steps of iterative approximation, Example 4.1, α_1=0.2, k_1=10, α_2=0.05, k_2=5.



convergence has been achieved. As a check the posterior probability ratio of the event that the 2, 3, 13 and 134 effects are active and observation y_{13} is bad, versus the event that effects 2 and 3 are active with no bad observations, was computed. The value of 17,186 indicates that there is much stronger evidence for the former event.

For the revised data with y_{13}=55.15, the probabilities $\{p_i\}$, assuming no bad values, are plotted in Figure 4.6. Tentatively identifying the effects 2, 3 and 13 as active the probabilities $\{q_j\}$ were computed and are plotted in Figure 4.12. The probability that y_{13} is bad is .37. Choosing Q greater than .37 ends the iteration. Choosing Q so that y_{13} is identified as a bad observation, the probabilities $\{p_i\}$ were recomputed conditional on that assumption and are plotted in Figure 4.13. Effects 2, 3, 13 and 134 all received fairly high posterior probabilities and can be identified as active. Recomputing the $\{q_j\}$ based on this new model gives the values plotted in Figure 4.14. The probability that observation y_{13} is bad is now close to one and any reasonable choice of Q results in convergence. Checking the posterior probability ratio of the event that 2, 3, 13 and 134 are active and y_{13} is bad, versus the event that 2, 3 and 13 are active and no observations are bad gives the value 47.9, thus giving more weight to the former combination.

Comparing the "exact" probabilities in Figures 4.4a-b for the original data and Figures 4.9a-b for the revised data, with the approximate probabilities in Figures 4.10-4.11 for the original data and Figures 4.13-4.14 for the revised data, the values are in reasonable agreement. While the actual numerical values are not as close as one might prefer, the inferences following from the computed probabilities agree quite

Figure 4.12 Plot of posterior probabilities $\{q_j\}$ that observations are bad, after one step of iterative approximation, Example 4.1 with y_{13}=55.15, α_1=0.2, k_1=10, α_2=0.05, k_2=5.

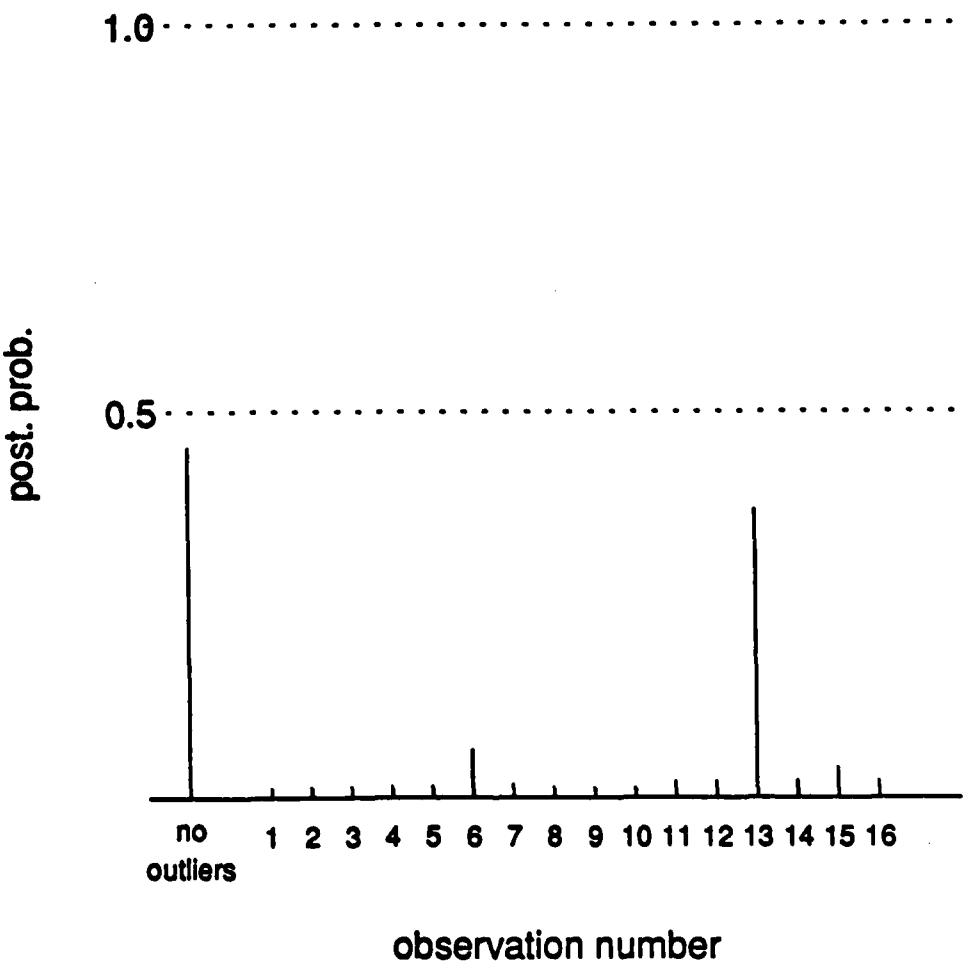


Figure 4.13 Plot of posterior probabilities $\{p_i\}$ that contrasts are active, after two steps of iterative approximation, Example 4.1 with y_{13}=55.15, α_1=0.2, k_1=10, α_2=0.05, k_2=5.

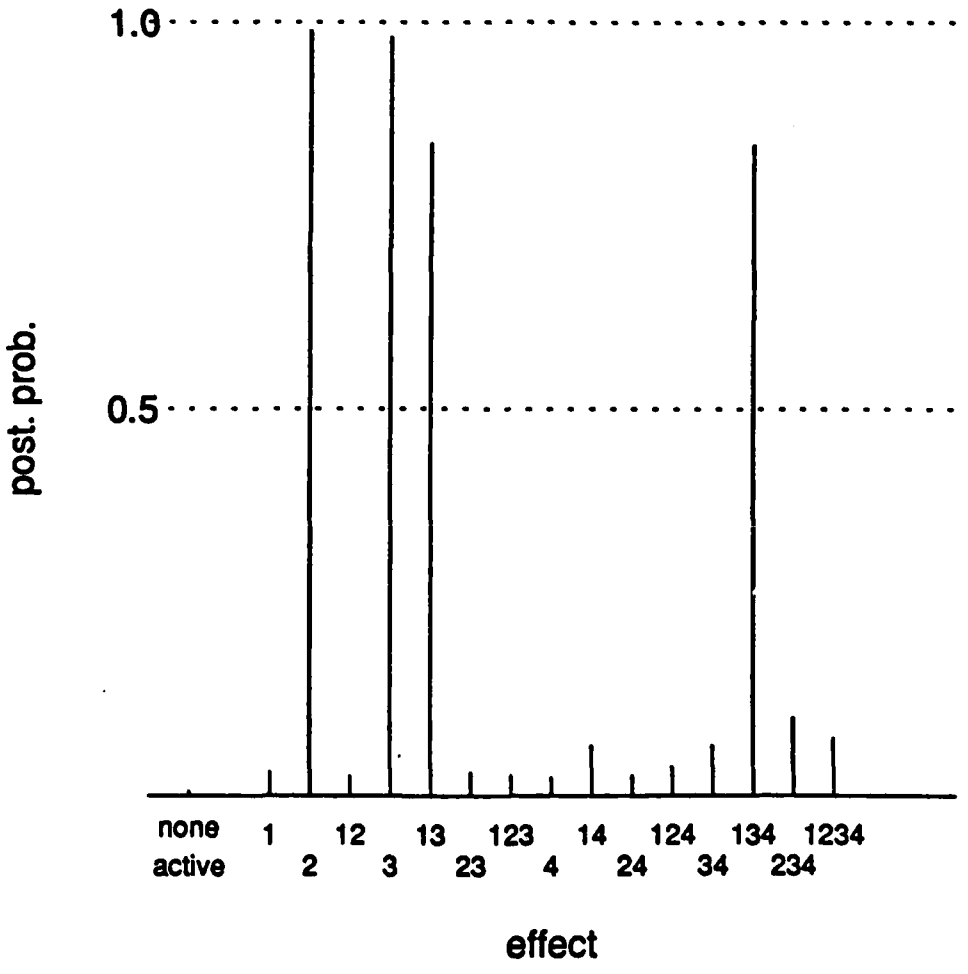
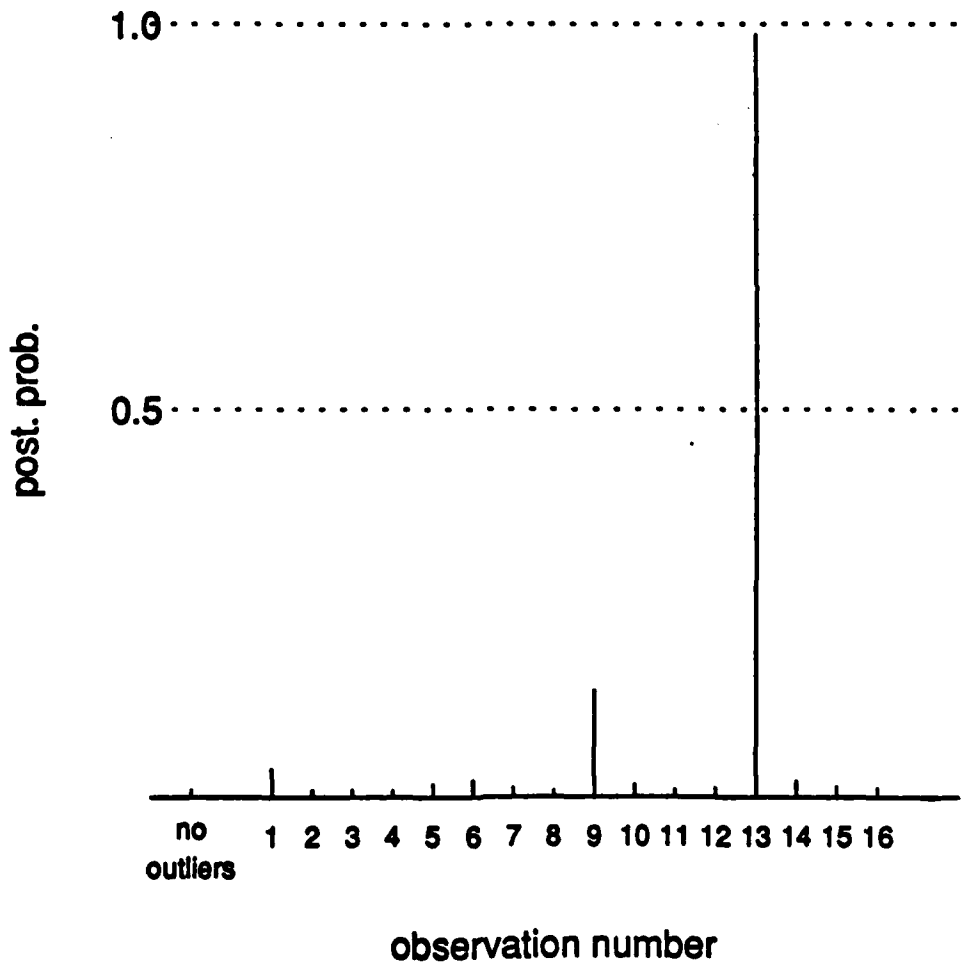


Figure 4.14 Plot of posterior probabilities $\{q_j\}$ that observations are bad, after two steps of iterative approximation, Example 4.1 with y_{13}=55.15, α_1=0.2, k_1=10, α_2=0.05, k_2=5.



well.

Reduction in computing time is a major benefit of this method of approximation. For the examples above, while the exact calculations required four hours computing time for each complete set of probabilities, the approximations required only 2-1/2 minutes. Both sets of calculations were done assuming six or fewer active contrasts and two or fewer bad observations. The lower computing requirements would allow this assumption to be relaxed when using the approximate method.

4.6. Posterior Distribution of τ

The posterior distribution of an effect τ_i for the Bayesian model not allowing for the possibility of bad observations, Chapter 2, was shown to be a mixture of 2^{n-2} t distributions with the same mean and different variances, together with mass $1-p_i$ at zero. Allowing for the possibility of bad observations, the posterior distribution of τ_i will have the same form, i.e.,

$$p(\tau_i \mid y) = (1-p_i) I[\tau_i=0] + \sum_{(r,c):i \text{ active}} p(\tau_i \mid a_{(r,c)}, y)\, p(a_{(r,c)} \mid y) \cdot \qquad (4.21)$$

However, the t densities in the mixture have different means as well as variances (see equations 4.12, 4.13). Following the approximation method of truncated Taylor series expansion, Chapter 2, three quadratic terms would be required rather than one because the expansion would be in terms of two variables, the mean and variance. The added complexity of such an approach does not seem practical given the computational limitations.

The posterior mean and variance of τ_i, given it is active, are

$$\hat{\tau}_i = E[\tau_i \mid \tau_i \text{ active}, \mathbf{y}] = \sum_{(r,c):i \text{ active}} E[\tau_i \mid a_{(r,c)}, \mathbf{y}] p(a_{(r,c)} \mid \mathbf{y}), \tag{4.22}$$

and

$$\text{Var}(\tau_i \mid \tau_i \text{ active}, \mathbf{y}) = V_1 + V_2 \tag{4.23}$$

where

$$V_1 = \sum_{(r,c):i \text{ active}} \text{Var}(\tau_i \mid a_{(r,c)}, \mathbf{y}) p(a_{(r,c)} \mid \mathbf{y}), \tag{4.24}$$

$$V_2 = \sum_{(r,c):i \text{ active}} \left[E[\tau_i \mid a_{(r,c)}, \mathbf{y}] - \hat{\tau}_i \right]^2 p(a_{(r,c)} \mid \mathbf{y}). \tag{4.25}$$

If the term V_2 in the posterior variance is negligible, then the methods of Chapter 2 will apply. That is, if the means of the t distributions in the mixture (4.21) do not vary significantly, the statistic CV defined by equation (2.41) can be used to construct a confidence interval for τ_i. Thus there would be two statistics to examine when deciding if using a single t interval is appropriate, the CV statistic and the proportion of the posterior variance due to variation in the mean,

$$\frac{V_2}{V_1 + V_2}.$$

At present the adequacy of this approach can not be explored due to the heavy computing requirements involved. It is hoped that future computing advances will allow this issue to be clarified.

4.7. Conclusions

The extended Bayesian model to allow for the possibility of bad observations has been demonstrated to lead to reasonable conclusions. Given the frequency of bad observations in industrial and other applications of factorial experiments, the analysis can be quite valuable. This is especially true for screening situations when a fixed model cannot be identified in advance, and the presence of bad observations may lead to erroneous identification of inert effects as active, or vice versa.

The computational limitations of the method are, of course, troublesome. The iterative approximation method described in this chapter is rationally motivated and leads to sensible results for the examples illustrated, but much more could be done from a numerical and algorithmic viewpoint.

CHAPTER 5

SUMMARY

Unreplicated factorial designs have been and still are a valuable tool in industrial experimentation, despite the fact they do not allow for the estimation of error variance usually obtained from repeat runs. Methods of analysis used in the past have depended more or less on an implicit assumption about the sparsity of real effects. If such assumptions are explicitly incorporated into the usual linear model employed for such experiments, inference about active and inert contrasts is more straightforward, and dependence of the inference on the prior assumptions is more easily assessed.

It is assumed that there is a prior probability α that each of the orthogonal contrasts is active, i.e., measures a real effect, and contrasts are active independently of one another. Assuming a normal prior distribution for the expected value of an active contrast, the posterior probability that a contrast is active can be computed. While computations of this sort generally require extensive computing time, an alternative Bayes factorization allows the posterior probabilities to be obtained by numerical integration at a considerable reduction in computing requirements. Dependence of the posterior probabilities on the choice of α and k, the inflation factor for an active contrast, can be measured by carrying out the calculations for different values of the parameters. Computation of the partial derivatives of the probabilities with respect to α and k will also give a measure of sensitivity. It was demonstrated in Chapter 2 that probabilities associated with in-between contrasts are the most sensitive to choice of

prior assumptions.

The posterior density of a true effect τ_i was shown to be a mixture of 2^{n-2} t densities along with discrete mass at zero of $1-p_i$, where p_i is the posterior probability that the effect τ_i is active. The continuous part of this distribution is often well approximated by a single t distribution, and by a Taylor series argument there is a coefficient of variation-like statistic CV which conveniently measures the closeness of the approximation.

Further assumptions about the size and relative frequency of main effects and interactions were incorporated into the model in Chapter 3. It was shown that the posterior probabilities that experimental factors are active combines prior assumptions, properties of the design and information in the data. Factors which can not be safely eliminated as inert due to the confounding pattern of the design will receive significant posterior probability in addition to those factors which are more obviously active.

A simulation study of sensitivity of the analysis to the assumption of normally distributed errors was carried out for two situations: with and without active effects present. Pseudo-random errors were generated by computer from three alternative distributions (one light-tailed, one heavy-tailed and one skew) as well as the normal. There was no evidence from the simulations that non-normal errors would affect the Bayesian analysis to any substantial extent. The posterior probabilities performed well in identifying active contrasts and factors for all four distributions tested.

The model was extended in Chapter 4 to allow for the possibility of bad observations. Observations were assumed to have inflated variance with prior probability α_2. Given this model, the posterior probability that a contrast (or factor) is active could be computed taking into account the possibility of bad observations, as well as the probability that a particular observation is bad, i.e., has inflated variance. It was shown that the approach of testing residuals for outliers after active contrasts are identified is sometimes inferior to the Bayesian model-based approach.

The extension to the possibility of bad observations greatly increases the computing requirements of the analysis, so that they are often unfeasibly high. An iterative analysis was proposed as an exploratory method rather than a numerical approximation. A method of approximating the posterior probabilities which possessed good numerical properties would be one area of future research.

References

Atkinson, A. C. (1978) Posterior Probabilities for Choosing a Regression Model. *Biometrika*, 65, 39-48.

Barnard, G. (1980) Discussion of Box (1980), Sampling and Bayes' Inference in Scientific Modelling and Robustness. *J. R. Statist. Soc. A*, 143, 404-406.

Barnett, M. K. and Mead, F. C. (1956) A 2^4 Factorial Experiment in Four Blocks of Eight: A Study in Radioactive Decontamination. *Appl. Statist.*, 5, 122-131.

Beckman, R. J. and Cook, R. D. (1983) Outlier...s. *Technometrics*, 25, 119-164.

Bennett, C. A. and Franklin, N. L. (1954) *Statistical Analysis in Chemistry and the Chemical Industry*. New York: John Wiley.

Birnbaum, Allan (1959) On the Analysis of Factorial Experiments Without Replication. *Technometrics*, 1, 343-357.

Box, George E. P. and Wilson, K. B. (1951) On the Experimental Attainment of Optimum Conditions. *J. R. Statist. Soc. B*, 13, 1-45.

Box, G. E. P. and Hunter, J. S. (1957) Multi-factor Experimental Designs for Exploring Response Surfaces. *Ann. Math. Statist.*, 28, 195-241.

Box, George E. P. and Hunter, J. Stuart (1961) The 2^{k-p} Fractional Factorial Designs. *Technometrics*, 3, 311-351, 449-458.

Box, George E. P. and Tiao, George C. (1968) A Bayesian Approach to Some Outlier Problems. *Biometrika*, 55, 119-129.

Box, G. E. P. and Tiao, G. C. (1973) *Bayesian Inference in Statistical Analysis*. Reading, Mass.: Addison-Wesley.

Box, George E. P. and Jenkins, Gwilym M. (1976) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Box, George E. P., Hunter, William G. and Hunter, J. Stuart (1978) *Statistics for Experimenters*. New York: John Wiley.

Box, George E. P. (1980) Sampling and Bayes' Inference in Scientific Modelling and Robustness. *J. R. Statist. Soc. A*, 143, 383-430.

Box, G. E. P. and Meyer, R. D. (1985) Analyzing Two-Level Fractional Factorial Experiments for Possible Dispersion Effects. TSR#2746, Mathematics Research Center, University of Wisconsin.

Box, G. E. P. and Draper, N. R. (1986) *Empirical Model-Building and Response Surfaces*. New York: John Wiley.

Chen, Gina and Box, George E. P. (1979) Further Study of Robustification via a Bayesian Approach. TSR#1998, Mathematics Research Center, University of Wisconsin.

Cochran, W. G. (1941) The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total. *Annals of Eugenics*, **11**, 47-52.

Cochran, W. G. and Cox, G. M. (1957) *Experimental Designs*. New York: John Wiley.

Daniel, Cuthbert (1959) Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments. *Technometrics*, **1**, 311-341.

Daniel, C. (1960) Locating Outliers in Factorial Experiments. *Technometrics*, **2**, 149-156.

Daniel, Cuthbert (1976) *Applications of Statistics to Industrial Experimentation*. New York: John Wiley.

Deming, W. Edwards (1982) *Quality, Productivity, and Competitive Position*. Cambridge, Mass.: MIT Center for Advanced Engineering Study.

Dixon, W. J. (1951) Ratios Involving Extreme Values. *Ann. Math. Statist.*, **22**, 68-78.

ed., O. L. Davies, (1954) *Design and Analysis of Industrial Experiments*. London: Oliver and Boyd.

Finney, D. J. (1945) The Fractional Replication of Factorial Arrangements. *Annals of Eugenics*, **12**, 291-301.

Fisher, R. A. (1966) *The Design of Experiments, 8th edition*. New York: Hafner.

Gentleman, J. F. and Wilk, M. B. (1975) Detecting Outliers in a Two-Way Table: I. Statistical Behaviour of Residuals. *Technometrics*, **17**, 1-14.

Gentleman, J. F. and Wilk, M. B. (1975) Detecting Outliers: II. Supplementing the Direct Analysis of Residuals. *Biometrics*, **31**, 387-410.

Goldsmith, P. L. and Boddy, R. (1973) Critical Analysis of Factorial Experiments and Orthogonal Fractions. *Appl. Statist.*, **22**, 141-160.

Holms, A. G. and Berrettoni, J. N. (1969) Chain-Pooling ANOVA for Two-Level Factorial Replication-Free Experiments. *Technometrics*, **11**, 725-746.

Jeffreys, H. (1932) An Alternative To the Rejection of Observations. *Proceedings of the Royal Society, London, A*, **137**, 78-87.

John, J. A. and Prescott, P. (1975) Critical Values of a Test to Detect Outliers in Factorial Experiments. *Appl. Statist.*, **24**, 56-59.

John, J. A. and Prescott, P. (1975) Estimating Missing Values in Experiments. *Appl. Statist.*, **24**, 190-192.

John, J. A. (1978) Outliers in Factorial Experiments. *Appl. Statist.*, **27**, 111-119.

John, J. A. and Draper, N. R. (1978) On Testing for Two or One Outliers in Two-Way Tables. *Technometrics*, **20**, 69-78.

Johnson, N. L. and Leone, F. C. (1964) *Statistics and Experimental Design, Volume 2*. New York: John Wiley.

Little, J. K. (1983) Analysis of the Scale-Contaminated Normal Model: Diagnostics and Robustness. Ph.D. Thesis, Department of Statistics, University of Wisconsin.

Plackett, R. L. and Burman, J. P. (1946) Design of Optimal Multifactorial Experiments. *Biometrika*, **23**, 305-325.

Rao, C. R. (1947) Factorial Experiments Derivable from Combinatorial Arrangements of Arrays. *JRSS*, **B9**, 128-140.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*. New York: John Wiley.

Stefansky, W. (1972) Rejecting Outliers in Factorial Designs. *Technometrics*, **14**, 469-479.

Taguchi, Genichi and Wu, Y. (1980) *Introduction to Off-line Quality Control*. Nagoya, Japan: Central Japan Quality Control Association.

Tippett, L. H. C. (1934) Applications of Statistical Methods to the Control of Quality in Industrial Production. *Manchester Statistical Society*, .

Tukey, J. W. (1960) A Survey of Sampling From Contaminated Distributions. Essays in Honor of Harold Hotelling, J. Olkin et al. (eds.), Stanford University Press.

Wilk, M. B., Gnanadesikan, R. and Freeny, A. E. (1963) Estimation of Error Variance From Smallest Ordered Contrasts. *J. Amer. Statist. Ass.*, 58, 152-160.

Zahn, Douglas A. (1975) Modifications of and Revised Critical Values for the Half-Normal Plot. *Technometrics*, 17, 189-200.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #2865	2. GOVT ACCESSION NO. AD-A160 972	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ANALYSIS OF FACTORIAL EXPERIMENTS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. Daniel Meyer		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041 DMS-8420968
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS (See Item 18 below)		12. REPORT DATE September 1985
		13. NUMBER OF PAGES 173
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, D. C. 20550

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Fractional factorial, unreplicated, active effects, Bayesian inference, outliers

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Fractional factorial designs have long been a key tool for the industrial statistician. They have received renewed attention recently due to the movement toward quality improvement sparked by the success of the Japanese in penetrating markets formerly dominated by western countries.

Fractional factorial designs are usually not replicated, so that it is not possible to estimate error variance in the usual way from repeat observations. Past methods of analysis have rested on an implicit hypothesis of

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

effect sparsity, that most of the estimated effects measure only noise. Formalization of this hypothesis leads to a Bayesian analysis in which the posterior probability that an effect is active can be computed. A similar approach can be employed to obtain the posterior probability that a particular experimental factor is active. These probabilities are readily interpreted by graphical means, and provide a straightforward method for identifying active contrasts and active factors. In addition, the model is extended to the situation where there are possible outliers in the original observations. The posterior probability that an effect is active can be computed taking into account the possibility of bad values, and the posterior probability that an observation is bad can be computed taking into account that the identity of active effects is unknown.

END

FILMED

12-85

DTIC