Detecting Active Effects in Unreplicated 2^k Factorial Experiments

– A Simulation Study –

by

Pieter Theo Pepler

Assignment submitted in partial fulfillment of the requirements for the degree

MASTER OF COMMERCE

in the Department of Statistics and Actuarial Science at the University of Stellenbosch

Supervisor: Dr. J.O. van Vuuren

November 2006

Contents

Chapter 1:
Introduction

1.1 Introduction

Experimentation is an integral component of modern research investigation, and the practicalities thereof always involve cost considerations. Factorial experiments provide a well-structured ingenious methodology to extract the most possible information from as few experimental runs as possible. For this very reason, unreplicated 2^k factorial experiments have been popular in statistical experimentation.

It is not without inherent difficulties, however: estimating which variables are the important ones (i.e. variable selection) proved to be more problematic than a superficial overview of the literature would necessarily suggest. This assignment is another attempt at evaluating some of the procedures proposed in past research on this problem, and building on previous work by suggesting some new alternatives.

The methodology and philosophy behind the use of unreplicated 2^k factorial experiments will be explained in this first chapter, as well as the formulation of the research problem being tackled in this assignment. Chapter 2 will briefly review the past work done on this problem, discussing a number of previously proposed methods. Some new proposals are made and illustrated in Chapter 3, and then compared to a few of the more well-known techniques by way of an extensive

simulation study discussed in Chapter 4. Chapter 5 concludes the assignment with a discussion of the results from the simulation study, as well as some suggestions for future research on this problem.

1.2 Background

Engineers study and develop processes in order to solve real-world problems. They use the scientific method to develop models to explain real phenomena, and they have to collect data to test such models (Vining 1998: 9). The engineering method involves gathering and analysing information about the processes under study, in order to improve them, and that is where statistical thinking and methodology come into play. Data collection, analysis, model building, and interpretation of models form a central part in engineering (Vining 1998: 1).

Statistical methodology is widely used throughout different stages of engineering processes. From the beginning stages of pilot plant operation, to process modelling, determining the important factors in processes, process optimization, and the eventual monitoring of processes, statistical thinking and philosophy has a fundamental role to fulfil. “Modern engineering statistics may be viewed as a symphony” (Vining 1998: 2).

The quality of the data obtained to study engineering processes is very important. No amount of statistical methodology or data manipulation can correct the inevitable biases (which can lead to wrong conclusions) resulting from poor sampling methods or taking an unrepresentative sample. Therefore statistical thinking has to be employed even in the initial stages of developing sampling plans and engineering experiments.

This paper will explore one facet of the engineering method, namely the data collection and analysis strategy known as experimental design. Specifically, attention will be given to experimental designs used in the initial stages of studying processes, and so-called screening experiments. Screening experiments attempt to distinguish the factors which influence a process from those which do not significantly influence it (Vining 1998: 19). It involves planning experiments in such a way as to use as little experimental runs as possible, while still gaining a wealth of information about the phenomena under study.

1.3 The use of 2^k factorial experiments

In studying the factors (process variables) influencing a process, one can go about it in a number of ways. One way is to conduct a number of single-factor experiments, studying only one factor per experiment. This method can work well if the goal is to obtain a broad view of the size and direction of the effects of the different factors, rather than finding the optimal combination of process variable levels to yield a maximum response (Cochran & Cox 1957: 150).

Another approach would be to conduct a factorial experiment, in which all the factors are included simultaneously. A factorial experiment has two main advantages over the single-factor approach: firstly, with a factorial experiment, each of the main effects can be estimated with the same precision as if the whole experiment was dedicated to that main effect alone. In other words, a factorial experiment studying two factors will need only half the number of experimental runs in order to estimate the main effects of the two factors with the same precision as two single-factor experiments. In general, for estimating the main effects of *n* factors with equal precision, the single-factor approach would require *n* times the number of experimental runs needed by a factorial experiment (Cochran & Cox 1957: 150-151). The factorial approach can therefore result in considerable savings in time, experimental materials, and money.

The second advantage of factorial experimentation is that factor interactions can be estimated. If the factors in a particular process are not independent, the simple effect of a factor may be significantly influenced by the specific combination other factors. The single-factor approach will not be able to pick up such interdependence among the process variables. In a factorial experiment, each of the factors is investigated at every combination of the other factors, enabling the data analyst to estimate the size, direction, and significance of interaction between the different factors (Cochran & Cox 1957: 151).

Screening experiments, in which the aim is to make an exploratory investigation of the magnitude and direction of the effects of the factors of the process under study, typically involve a large number of factors. If each of these factors (process variables) has to be studied over a range of levels, the number of experimental runs required, to estimate all main effects and interaction effects, can quickly become prohibitively large (Hinklemann & Kempthorne 1994: 352). For example, if only 20 factors has to

be investigated, each at just three levels of operation, the number of runs to have one complete replicate of the experiment is equal to 3^{20} = 3 486 784 401!

Screening experiments are also mostly used as the starting point in a sequential experimentation process, to first screen out all the insignificant factors. Thereafter, more intensive experimentation can be performed on the remaining factors, to find the combination of variable levels which optimizes the process under study. In the light of this philosophy of experimentation, it would seem appropriate to use the least amount of experimental runs in the beginning stages (i.e. with the screening experiments) of the sequential experimentation process, and "save" resources for the later ("optimization") stages, when more observations will be needed to fit more complex models in the region of the process optimum (Vining 1998: 353).

One way to overcome this problem is to use only a subset of the possible treatment combinations from a full factorial experiment. This is called a fractional factorial experiment (Hinklemann & Kempthorne 1994: 352). Depending on the subset of treatment combinations chosen for the fractional factorial, it limits the estimation of certain of the main and/or interaction effects, though.

Another way is to perform the factorial experiment on all the factors of interest, say *k* of them, but restricting the number of levels which each factor can take on to two. The two levels of each factor should then be chosen so that it covers a reasonable range within practical operating conditions. Such an experiment is known as a 2^k factorial design. The 2^k factorial design is often used in industrial experimentation, but because it restricts the number of levels which each factor can take to only two, it is most useful in exploratory (screening) experiments. If a factor can take on more than the chosen two levels, the nature of the main effects and interactions cannot be studied in much detail. With quantitative factors, for example, a 2^k design will enable us to study only linear relationships between the factors (Hinklemann & Kempthorne 1994: 352-353).

Another disadvantage of these designs is the exponentially increasing size and complexity of a single factorial experiment, as the number of factors gets larger. Efficiency of factorial experimentation is the greatest when the factors are numerous, though (Cochran & Cox 1957: 152).

Analysis and summary of the results from a factorial experiment may also take more time and expertise than when using the single-factor experimentation approach (Cochran & Cox 1957: 153), but this should not be seen as a significant obstacle. The advantages of factorial experimentation still outweigh the disadvantages.

1.4 Modelling the 2^k factorial experiment

When a factorial experiment is run in an effort to acquire information about the role of the different factors influencing a process, the data need to be analysed in order to make sense of it. Usually some sort of model will be fit to the experimental data. The parameter estimates for the fitted model may then be interpreted to make inference about the size and significance of the factors involved in the process.

The simplest possible model construction for the data obtained from a 2^k factorial experiment would be a model of the form:

$$y_i = \beta + e_i \qquad (1.1)$$

where $i = 1, 2, \ldots, n$, and each individual response (y_i) is modelled by the overall mean (β), plus some individual error term (e_i). The single parameter β is estimated by taking the mean of the values of the response variable for the *k* runs from the factorial experiment, and it represents the deterministic component in the model. The individual error terms $\{e_i\}$ are the stochastic (or random) part of the model, and can be calculated as the difference between each observed value of the response variable (for each of the *i* experimental runs) and the overall mean response β (Box & Draper 1987: p47-50).

A better approach, however, would be to make use of the fact that certain factors in the process will possibly have an effect on the outcome of interest. If these effects are real and consistent, one should be able to model them using a deterministic regression model. In such a model, the predicted response will have been adjusted for the effects of the different process factors, and should be closer to the real observed value. In other words, the stochastic (or random) error component in the model will be decreased, as one would be able to account for more of the variation in

the response variable due to the effects of the different active process factors being included in the model. This is known as an *effects model* (Montgomery 2001: 64).

1.4.1 Factors, levels, and design variables

In 2^k factorial experiments, each process factor is investigated at only two nominally specified levels. Every possible treatment combination of the factors (each at only the two specified levels) is then tested – as a whole, this makes up one complete replicate of the factorial design (Vining 1998: 359).

Usually, since each factor only has two levels, one is arbitrarily labelled the "low" (or "-1") level, and the other then the "high" (or "+1") level (Vining 1998: 359). Naming the factor levels as such can be interpreted in a number of ways, according to the nature (or more specifically, the measurement level) of a specific factor. For factors which are measured on a numerical interval scale, the low and high settings will denote certain prespecified levels of that factor. These levels are usually chosen as to fall in or close to the range of the normal operating conditions of the process under investigation. If one, for example, needs to examine the effect of temperature in a certain process, the low and high levels of temperature can be chosen as 120 and 155 degrees Celsius, respectively.

For qualitative factors, low and high can be used to indicate the absence or presence of a factor. For example, in a agricultural research experiment studying the effect of fertilizers on crop yield, the high level of a nitrogen factor can indicate that the specific plot (experimental run) have been treated with nitrogen, whereas nitrogen at the low level will then indicate that the plot have not been treated with nitrogen. The high and low level terms can also be used for two different categories of a categorical variable, for example, having a cooling system factor, with its low level indicating a water-cooled and its high level indicating an air-cooled system (Milliken & Johnson 1989: 93).

The "+1" and "-1" indicators are called the *design variables* (Vining 1998: 359), and are used to ease the setting up of the design matrix for the factorial experiment. A complete replicate of a 2^2 factorial design simply consists of all possible combinations of the two factors at their respective two levels. In tabulated form, such a design can be represented as in *Table 1.1* below, or geometrically as shown in *Figure 1.1*.

x_1	x_2
-1	-1
1	-1
-1	1
1	1

Source: Vining 1998: 360

Table 1.1: The design matrix for an unreplicated 2^2 factorial design, making use of design variables.

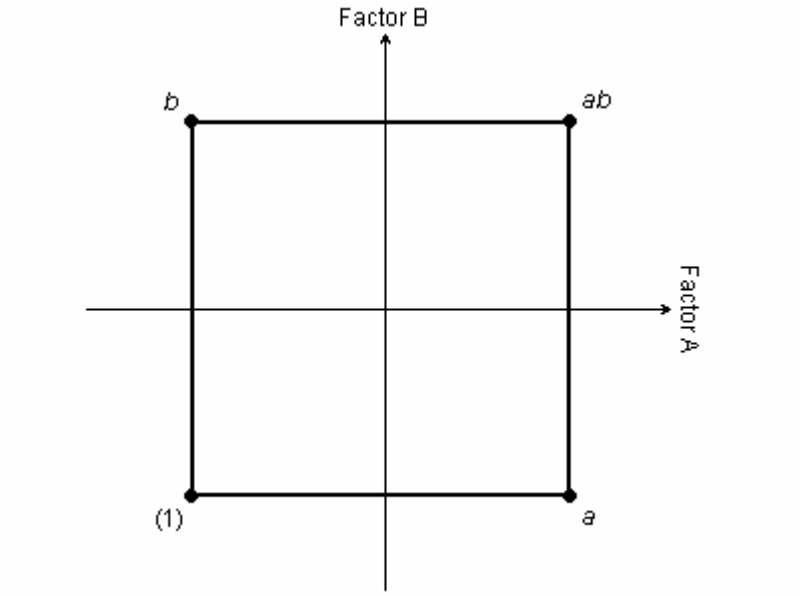

Figure 1.1: Geometric representation of a 2^2 factorial experiment.

As can be seen from *Figure 1.1*, the different treatment combinations can also be denoted as *(1)*, *a*, *b*, and *ab*, where any specific letter indicates that the corresponding effect is at its high level, e.g. *a* is the treatment combination where effect *A* is at its high level and *B* at its low level. *(1)* is the treatment combination where all effects are at their low levels (Montgomery 2001: 219-220).

Another useful property of defining the original factor variables in terms of design variables is that fitting a regression model to a complete factorial design results in a mean-corrected model. Suppose one has a simple linear regression model of the form

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

with $i = 1, 2, \ldots, n$. The standard estimate for the intercept term is then

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

An alternative to this model is

$$y_i = \beta_{*0} + \beta_1 (x_i - \bar{x}) + e_i$$

where the independent variable is adjusted by subtracting its mean. Using this mean-corrected form to model a complete replicate of a factorial design, the independent (predictor) variable has the property $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$. The mean of the newly defined independent variable $(x_i - \bar{x})$ therefore is zero (one can see this by inspecting *Table 1.1*: the mean for each of the two columns is equal to zero) and the estimate for the intercept term then reduces to

$$\hat{\beta}_{*0} = \bar{y}$$

Using a mean-corrected model therefore leads to a simple and easily interpretable estimate of the intercept term with the intuitive appeal that it is the estimated mean of the response variable. It also leads to the fact that $\bar{y}$ and $\hat{\beta}_1$ are independent, which simplifies the computation of variances for the regression line estimates (Christensen 1996: 178).

For $k = 2$ factors, an effects model (with a term for interaction) can now be specified as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i \qquad (1.2)$$

where $i = 1, 2, \ldots, n$ denotes the ith observation (experimental run). This is a first-order linear regression model: first-order because none of the dependent variables are raised to a power of more than one, and linear because it is linear in its parameters. Given that x_1 and x_2 are centred around zero (i.e. one is using design variables, and not the "natural variables"), the intercept term, β_0, is the estimated overall mean response. β_1 and β_2 are the partial effects due to the first and second factors, respectively, on the outcome of the process. Any variation not modelled by the deterministic part of the model, will be encapsulated in the error term e_i. This will include random variation in the process, but also any systematic variation which is not due to (and therefore not modelled by) the two factors included in the model.

1.4.2 Interaction

When the effect of one of the process factors depends on the specific level used for one (or more) of the other factors, there is said to be interaction between the factors. This concept is most easily grasped by looking at an interaction plot, which is a graphical display of the interaction effect (Vining 1998: 363).

For a 2^k factorial design, an interaction plot consists of plotting the means of one factor at its two levels while keeping the level of another factor constant. Given the levels of, say, factor B, the means for factor A are plotted at the low and high levels of A, and a graph such as the one shown in *Figure 1.2* might result (Vining 1998: 364).

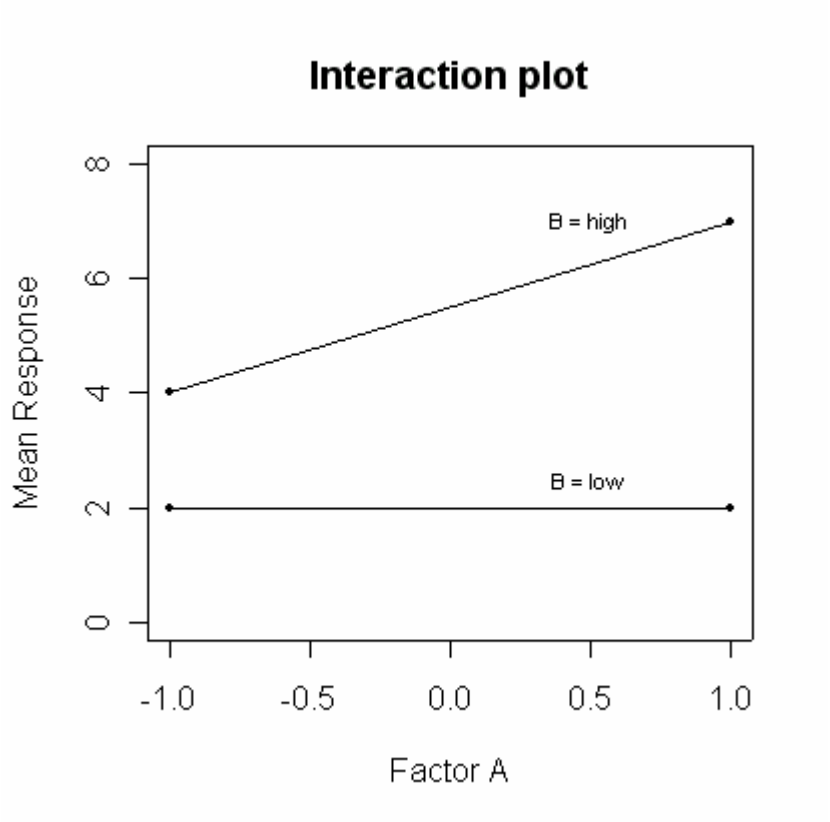

Figure 1.2: Interaction plot of factors A and B. The effect of A clearly depends on the level of B, indicating interaction between these two factors.

When B is at its high level, A has a positive effect on the response variable of the process under study. At the low level of B, however, adjusting A from its low to its high level has no effect on the response. The effect of A therefore depends on the specific level of B, which indicates that there is an interaction effect between factors A and B.

1.4.3 Model assumptions

The regression model given by (1.2) can be rewritten as follows:

$$y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i \qquad (1.3)$$

where the x_0 variable will just be equal to one for all $i = 1, 2, \ldots, n$ observations.

The following data matrix and vectors will now be defined (Searle 1971: 79):

$$X = \begin{bmatrix} x_{10} & x_{11} & x_{12} \\ x_{20} & x_{21} & x_{22} \\ . & . & . \\ . & . & . \\ . & . & . \\ x_{n0} & x_{n1} & x_{n2} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

In matrix form, the complete set of equations represented by (1.3) is then given by

$$Y = X\beta + \varepsilon \tag{1.4}$$

Using a model of the form of (1.3) requires making certain assumptions about the distributional properties of error term (Searle 1971: 80). Firstly, one assumes that the mean (expected value) of the errors is zero, that is

$$E(\varepsilon) = 0$$

This means that

$$E(Y) = X\beta$$

which one gets from taking the expected value of (1.4). The expected (estimated) value for any y_i is therefore given simply by multiplying its observation vector X_i with the coefficient vector β.

These results can straightforwardly be extended to a model with *k* variables (with all included interactions also counted as variables) (Searle 1971: 79). For a model with *k* variables

$$X = \begin{bmatrix} x_{10} & x_{11} & . & . & . & x_{1k} \\ x_{20} & x_{21} & . & . & . & x_{2k} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ x_{n0} & x_{n1} & . & . & . & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}$$

with Y and ε defined as before. (1.4) also remains unchanged as the representation of the model in matrix form, as long as the number of variables k does not exceed the number of observations n. When $k \geq n$, a solution can be found so that $Y = X\beta$ exactly, which means that $\varepsilon = 0$ and there is no estimation problem (Searle 1971: 79), and consequently no estimate of the error variance.

A further assumption is that the errors are identically and independently distributed, with an unknown but constant variance

$$Cov(\varepsilon) = \sigma^2 I_n$$

For hypothesis testing and calculation of confidence intervals, one additionally needs to make the assumption that the errors are normally distributed (Searle 1971: 146)

$$\varepsilon \sim N(0,\ \sigma^2 I_n)$$

Then, because of this assumption

$$Y \sim N\left(X\beta,\ \sigma^2 I_n\right)$$

i.e. the estimated response, Y, is also normally distributed, and, since the estimated regression coefficients are linear functions of Y, they are also normally distributed (Searle 1971: 174).

1.4.4 Least squares and the normal equations

In statistical model building, the aim is to find a model for which the estimated values for the response variable are as close as possible to the actual observed values. In other words, one wants to minimize the error in estimation. Because $E(\varepsilon) = 0$, instead of minimising the errors directly, one has to minimise the squared errors. The estimation error is the difference between the observed value of y_i and the estimated value for y_i, i.e. $E(y_i)$.

The squared errors is therefore given by

$$\begin{aligned}\varepsilon'\varepsilon &= \left[Y - E(Y)\right]' \left[Y - E(Y)\right] \\ &= (Y - X\beta)'(Y - X\beta) \\ &= Y'Y - 2\beta' X'Y + \beta' X'X\beta\end{aligned}$$

Because both Y and X are known, minimising the estimation error constitutes finding the estimator $\hat{\beta}$ such that $\varepsilon'\varepsilon$ is at its minimum. This is done by differentiating $\varepsilon'\varepsilon$ with respect to β and then equating $\partial(\varepsilon'\varepsilon)/\partial\beta$ to zero which gives

$$X'X\beta = X'Y$$

These are known as the *normal equations* (Searle 1971: 80). By making β the subject of the equation, one arrives at the unique estimator for β

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y \qquad (1.5)$$

Let the design matrix X be written in terms of design variables in standard order, such as (for a single complete replicate of a 2^3 design)

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

or, in a more general form

$$X = \begin{bmatrix} \underline{1} & \underline{c}_1 & \underline{c}_2 & \underline{c}_3 & \underline{c}_{12} & \underline{c}_{13} & \underline{c}_{23} & \underline{c}_{123} \end{bmatrix}$$

where each $\underline{c}_i$ is a column vector representing a contrast for the specific effect.

Alternatively

$$X = \begin{bmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \\ \underline{r}_4 \\ \underline{r}_5 \\ \underline{r}_6 \\ \underline{r}_7 \\ \underline{r}_8 \end{bmatrix}$$

where $\underline{r}_i$ then represents the ith row vector of the design matrix.

The result is an *orthogonal design*, for the cross products of any of the row and/or column vectors are equal to zero (Montgomery 2001: 404). Because of its orthogonality, the cross product of X is a diagonal matrix:

$$X'X = m2^k I_{2^k}$$

which, for a single replicate ($\therefore m = 1$), reduces to

$$X'X = 2^k I_{2^k}$$

Computation of the inverse of $X'X$ is therefore greatly simplified, given by

$$(X'X)^{-1} = \frac{1}{m2^k} I_{2^k}$$

which again, for a single replicate, reduces to

$$(X'X)^{-1} = \frac{1}{2^k} I_{2^k}$$

Furthermore

$$X'Y = \begin{bmatrix} \underline{1}'Y \\ \underline{c}_1'Y \\ \underline{c}_2'Y \\ \underline{c}_3'Y \\ \underline{c}_{12}'Y \\ \underline{c}_{13}'Y \\ \underline{c}_{23}'Y \\ \underline{c}_{123}'Y \end{bmatrix}$$

and then

$$\hat{\beta} = (X'X)^{-1} X'Y = \frac{1}{m2^k} \begin{bmatrix} \underline{1}'Y \\ \underline{c}_1'Y \\ . \\ . \\ . \\ \underline{c}_{123}'Y \end{bmatrix}$$

From (1.4), the predicted response is then given by

$$\hat{Y} = X\hat{\beta} \qquad (1.6)$$

and

$$\hat{y}_i = x_i'\hat{\beta} = \underline{r}_i\hat{\beta}$$

is the predicted response for any specific experimental run (Montgomery 2001: 396). The vector of residuals is simply the difference between the actual observations and the fitted values:

$$\varepsilon = Y - \hat{Y}$$

To estimate σ^2, the *mean square error* (*MSE*) is calculated as (Montgomery 2001: 397)

$$\hat{\sigma}^2 = \frac{SSE}{n-p}$$

where the *sum of squares of the residuals* (*SSE*) is given by

$$\begin{aligned} SSE &= \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \\ &= \sum_{i=1}^{n} e_i^2 \\ &= \varepsilon' \varepsilon \end{aligned}$$

1.4.5 Variance of the estimators

The variance of the estimated regression coefficients is (Montgomery 2001: 397)

$$Cov\left(\hat{\beta}\right) = \sigma^2 \left(X'X \right)^{-1} = \frac{\sigma^2}{m2^k} I_{2^k}$$

from which can be seen that the $\hat{\beta}_j$'s are uncorrelated (all the off-diagonal elements of the covariance matrix equals zero, i.e. $Cov\left(\hat{\beta}_i, \hat{\beta}_j\right) = 0$) (Montgomery 2001: 404). The variances of the $\hat{\beta}_j$'s are also all equal.

Working from (1.5) and (1.6)

$$\begin{aligned} \hat{Y} &= X\hat{\beta} \\ &= X\left(X'X \right)^{-1} X'Y \\ &= HY \end{aligned}$$

one can see that the projection (or "hat") matrix is defined as $H = X\left(X'X \right)^{-1} X'$ (Montgomery 2001: 417).

For the 2^k factorial design, the hat matrix can then be simplified as follows:

$$
\begin{aligned}
H &= X\left(X'X\right)^{-1}X' \\
&= X\left(\frac{1}{m2^k}\right)IX' \\
&= \left(\frac{1}{m2^k}\right)\begin{bmatrix} \underline{r}_1 \\ . \\ . \\ . \\ \underline{r}_{m2^k} \end{bmatrix}\begin{bmatrix} \underline{r}_1' & . & . & . & \underline{r}_{m2^k}' \end{bmatrix} \\
&= \left(\frac{1}{m2^k}\right)\begin{bmatrix} \underline{r}_1\underline{r}_1' & & 0 \\ & \ddots & \\ 0 & & \underline{r}_{m2^k}\underline{r}_{m2^k}' \end{bmatrix}
\end{aligned}
$$

Because of using design variables to construct the design matrix (which results in the fact that $\underline{r}_i\underline{r}_i' = m2^k$),

$$
\begin{aligned}
H &= \left(\frac{1}{m2^k}\right)\begin{bmatrix} m2^k & & 0 \\ & \ddots & \\ 0 & & m2^k \end{bmatrix} \\
&= \frac{1}{m}I_{m2^k}
\end{aligned}
$$

Therefore the leverage is constant and equal to $1/m$ for all observations (responses from the experimental runs). For a single replicate of a factorial design, the leverage of any observation will then of course simply be equal to one.

The variance of the predicted response is then given by

$$Var\left(\hat{y}_i\right) = h_{ii}\hat{\sigma}^2 = \frac{\hat{\sigma}^2}{m}$$

where h_{ii} is the ith diagonal element of the hat matrix. Therefore, for a single replicate,

$$Var\left(\hat{y}_i\right) = \frac{\hat{\sigma}^2}{1} = \hat{\sigma}^2$$

The covariance of the errors is given by (Montgomery 2001: 417)

$$Cov(\varepsilon) = \hat{\sigma}^2 (I - H)$$

from which the variance of an individual error then follows as

$$Var(\hat{e}_i) = \hat{\sigma}^2 (1 - h_{ii})$$
$$= \hat{\sigma}^2 \left(\frac{m-1}{m} \right)$$

One can see that for a single replicate of a factorial experiment, the second factor in the expression will become zero, and there will be no error variance (and consequently no error in estimation, for $E(\hat{e}_i) = 0$). The result is that all residuals will be zero, which is problematic for a number of reasons which will be discussed next.

1.4.6 Inference

For a regression analysis with *p* regressors (including the intercept), the entries for an *Analysis of Variance (ANOVA)* table (corrected for the mean) are calculated as follows (Searle 1971: 102):

ANOVA for regression

Source of variation	df	Sum of Squares	Mean Square	*F*-statistic
Regression	$p-1$	$SSR = \hat{\beta}' X'Y$	$MSR = \frac{SSR}{p-1}$	$F = \frac{MSR}{MSE}$
Error	$n-p$	$SSE = Y'Y - n\bar{y}^2 - \hat{\beta}' X'Y$	$MSE = \frac{SSE}{n-p}$	
Total	$n-1$	$SST = Y'Y - n\bar{y}^2$		

where n is the total number of observations. For a single replicate of a 2^k factorial experiment, one would have a total of 2^k observations/experimental runs (i.e. $n = 2^k$). However, if the *k* main factors and all their interactions are included in the regression model (which is usually done in the beginning of a study, as it is not yet known which

effects are important), the number of model parameters then also equals 2^k (i.e. $p = 2^k$). The ANOVA calculations for a situation such as this will look as follows:

ANOVA for a 2^k factorial design with one replicate

Source of variation	df	Sum of Squares	Mean Square	*F*-statistic
Regression	$p-1$	$SSR = \hat{\beta}' X'Y = Y'Y$	$MSR = \frac{SSR}{p-1}$	$F = ?$
Error	0	$SSE = -n\bar{y}^2$	$MSE = ?$	
Total	$n-1$	$SST = Y'Y - n\bar{y}^2$		

One can see that when $n = p$, there are no degrees of freedom left to estimate $\hat{\sigma}^2$ (the mean square error). Consequently the *F*-statistic for regression cannot be calculated, and one cannot make any conclusions about the statistical significance of the model effects (Vining 1998: 375).

Also, because the model then fits the data perfectly, $SSR = \hat{\beta}' X'Y = Y'Y$, and the *SSE* reduces to $-n\bar{y}^2$ (which equals zero for the mean adjusted model).

1.4.7 Regression diagnostics

After a regression model has been fit to the data, diagnostic techniques are employed to check the validity of the model assumptions, and consequently how well the model fits the data. One commonly used diagnostic measure is the *student residual* r_i, given by

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2\left(\frac{m-1}{m}\right)}}$$

When only a single replicate of the factorial design is run, i.e. $m = 1$, the denominator in the above expression becomes zero, and one therefore cannot calculate the student residual. Unfortunately this is not the only diagnostic procedure which fails

when using a single replicate – consider the *predicted error sum of squares (PRESS)* statistic:

$$PRESS = \sum_{n} e_{-i}^2$$

where $\{e_{-i}\}$ are the residuals resulting from fitting a regression model using all the data except the ith observation. Seeing as

$$e_{-i} = \frac{e_i}{1-h_i} = \frac{e_i}{1-\frac{1}{m}} = \frac{e_i}{\left(\frac{m-1}{m}\right)} = \left(\frac{m}{m-1}\right) e_i$$

the *PRESS* statistic then becomes

$$PRESS = \sum_{i=1}^{m2^k} \left(\frac{m}{m-1}\right)^2 e_i^2$$
$$= \left(\frac{m}{m-1}\right)^2 \sum_{i=1}^{m2^k} e_i^2$$

and once again the same problem presents itself in that, when $m=1$, the denominator of the factor $\left(\frac{m}{m-1}\right)^2$ becomes zero, and the *PRESS* statistic cannot be calculated.

For *Cook's statistic*, given by

$$Cook = \frac{r_i^2}{2^k} \frac{\left(\frac{1}{m}\right)}{\left(\frac{m-1}{m}\right)} = \frac{r_i^2}{2^k}\left(\frac{1}{m-1}\right) = \frac{r_i^2}{2^k\,(m-1)}$$

the same problem arises when only a single replicate of the experiment is run.

1.5 The problem

An *unreplicated factorial experiment* presents some problems. It has become clear in the preceding sections that one cannot simply proceed to fit a full regression model (all main effects and interactions) to the data from a single replicate of a 2^k factorial experiment, as the analysis cannot be done in the usual way.

Because there is only one replicate, there are no degrees of freedom left to estimate the internal error in the process (Montgomery 2001: 245). Consequently the usual *F*-statistics cannot be calculated to test the significance of the factor effects and interactions involved in the process. One therefore cannot draw inference about which factors has an effect on the response variable.

The usual regression diagnostic measures also fail with unreplicated factorial designs. As has been shown, none of the studentized residuals, *PRESS*, or *Cook's statistic* can be calculated for an unreplicated experiment.

To overcome this problem, there are two options open: either add more experimental runs (replicates, centre runs, etc.), which can be prohibitively expensive in the industrial environment where experimentation often involves large amounts of funds, or find a way to detect the real significant effects without having to augment the design in the aforementioned way.

1.6 Possible solutions

Although this problem had not yet been completely resolved, a number of helpful attempts had been proposed in the literature. Many of the proposed solutions appeal to the *sparsity of effects principle*, which is the assumption that most processes are dominated by some main and lower-order interaction effects, and that the higher-order interactions will either be nonexistent or negligibly small (Montgomery 2001: 245). By ignoring the higher-order interactions, and fitting a smaller model, degrees of freedom are "freed up" to obtain an estimate of the pure process variance.

In large processes, one can probably quite safely assume a sparsity of effects, but in smaller process involving only a few factors, the higher-order interaction effects may

very well be significant and cannot just be left out. “[O]ccasionally real high-order interactions occur. The use of an error mean square obtained by pooling high-order interactions is inappropriate in these cases” (Montgomery 2001: 246).

One of the earliest procedures which depends heavily on the sparsity of effects assumption, is Daniel’s suggestion of examining a normal probability plot of the estimated effects. “The effects that are negligible are normally distributed, with mean zero and variance σ^2 and will tend to fall along a straight line on this plot, whereas significant effects will have nonzero means and will not lie along the straight line” (Montgomery 2001: 246). This is sometimes referred to as the “fat pencil” rule (Vining 1998: 375).

There is however substantial subjectivity on the part of the analyst as to what constitutes effects lying on the line, and prompted the search for more objective procedures. The next chapter will provide an overview of several of the more effective methods which have been proposed over the past couple of decades.

Chapter 2:
Literature Review

2.1 Introduction

A sizable amount of research on the detection of active effects in unreplicated 2^k factorial experiments has amassed through the years. Some of the more well-known procedures will be briefly summarized in this chapter, outlining the methods and relating some findings made by other researchers.

With regards to categorizing the different procedures, the general groupings made by Hamada & Balakrishnan (1998: 15) will be followed. A distinction will be made between directed and composite methods, where the first refers to procedures which test the individual effects directly, while the latter test all the effects as a group. Within these categories, the methods will also be classified as sequential or non-sequential. For sequential methods, computation is done in a number of stages, using only the remaining contrasts at each stage and recalculating the critical values if needed.

A few of the proposed procedures can also be classified as hybrid methods, containing elements of more than one of the abovementioned categories. These methods will be discussed lastly.

2.2 Sequential directed methods

2.2.1 Daniel (1959)(2)

Daniel suggested a standardized probability plot of the unsigned contrasts, which , combined with guardrails, results in an objective graphical procedure for identifying active effects. The unsigned (absolute value) contrasts are standardized by dividing it by the unsigned contrast of which the unsigned order statistic is closest to the 0.683 percentile. This produces statistics called *modulus ratios*. The 0.683 percentile is used because "the 0.683 percentile of the half-normal distribution is equal to τ, and suggests an estimate for the contrast standard error τ when all the effects are inert" (Hamada & Balakrishnan 1998: 5).

In a 2^4 factorial screening experiment, one would typically want to estimate $k = 15$ effects, and because

$$0.683 \times 16 = 10.928 \approx 11$$

the unsigned contrasts will be standardized by $|c|_{(11)}$ to give the *modulus ratios*

$$\frac{|c|_{(i)}}{|c|_{(11)}}$$

where $|c|_{(i)}$ is the absolute value of the *i*th effect contrast. These modulus ratios are then plotted, together with their corresponding guardrails based on the distribution of $|c|_{(i)k}$. Any standardized contrasts which exceed their guardrails are identified as active effects. Note however "that only about 25% of the largest unsigned contrasts can be tested sequentially starting with the largest" (Hamada & Balakrishnan 1998: 5). One therefore have to make quite a strong sparsity of effects assumption when using this procedure.

"Birnbaum … showed that it is the most powerful test when there is only one active effect", which agrees with Hamada & Balakrishnan's findings that it performs better for small number of active effects (1998: 5, 22). They concluded it to be one of the better directed methods which work well for small number of active effects (Hamada & Balakrishnan 1998: 22).

2.2.2 Holms & Berrettoni (1969)

Holms & Berrettoni proposed a *chain-pooling* method which sequentially compares the largest standardized effect mean squares to the smallest mean squares. Assuming a sparsity of effects, the procedure is started by pooling the *m* (possibly equal to one) smallest effect mean squares and calculating the ratio

$$\frac{(m+1)M_{(m+1)}}{\sum_{i=1}^{m} M_{(i)}}$$

where $M_{(i)}$ is the *i*th ordered effect mean square. The *p*-value for this ratio is then compared to α_{pool} (e.g. 0.25) to determine whether the next largest mean square $M_{(m+1)}$ should be pooled or not. "Pooling is stopped once the *p*-value falls below α_{pool}" (Hamada & Balakrishnan 1998: 5).

The effects of which the *p*-value for the ratio

$$\frac{jM_{(l)}}{\sum_{i=1}^{j-1} M_{(i)} + M_{(l)}}$$

(where the $j-1$ smallest mean squares are pooled) falls below α_{final} are then deemed active. Critical values based on all inert effects for $k = 15$ are given in *Table 1* of Holms & Berrettoni (Hamada & Balakrishnan 1998: 5).

This method is therefore somewhat subjective, as a strategy is defined by the specific choice of *m*, α_{pool} and α_{final} (Hamada & Balakrishnan 1998: 5).

2.2.3 Seheult & Tukey (1982)

Seheult & Tukey proposed a method in which they consider a synthetic batch of the effect contrasts. From this batch, they then identify the outliers as the contrasts corresponding to active effects.

More specifically, the procedure consists of taking all the contrasts with both signs (i.e. the positive and negative of each contrast) and zero to form a group of $2^{k+1}-1$ items. The *interquartile range (IQR)* is then calculated, and the outliers are considered to be all contrasts which lie outside the interval

$$\left(Q_1 - 1.5IQR\ ;\ Q_3 + 1.5IQR\right)$$

where Q_1 and Q_3 are the first and third quartiles of the synthetic batch, respectively. Under the null hypothesis assumption that the contrasts are normally distributed with mean zero, the probability of exceeding these bounds is only 0.007 (Hamada & Balakrishnan 1998: 6).

This procedure can then be used iteratively by removing the largest contrast (both the positive and negative ones) from the batch if it falls outside the interval bounds, and applying the procedure to the remaining contrasts in the next iteration (Hamada & Balakrishnan 1998: 6). The process is stopped once no more outliers are found.

Hamada & Balakrishnan found this method to be "quite competitive with the best of the other methods for [a number of active effects] up to four and … slightly worse for six active effects" (Hamada & Balakrishnan 1998: 22).

2.2.4 Johnson & Tukey (1987)

In an approach very similar to Daniel's (1959), Johnson & Tukey proposed a *ratio-to-scale* method which employs *display ratios*. The display ratio for the *i*th contrast is given by

$$\frac{\left|c\right|_{(i)}}{\left|\tilde{z}\right|_{(i)k}}$$

where $|c|_{(i)}$ is the *i*th unsigned contrast, and $|\tilde{z}|_{(i)k}$ is the median of the half-normal *i*th order statistic in a sample of size *k*. These display ratios can be plotted to visually compare them to each other, and in general the plot will form a natural reference line of which the height is an estimate of τ.

Their *ratio-to-scale* statistics are then computed as

$$\text{ratio-to-scale} = \frac{\text{display ratio}}{\text{median display ratio}}$$

In their *Table 12*, they provided critical values for the *i*th maximum ratio-to-scale statistic for a sample of size *i*.

Johnson & Tukey suggested a sequential approach, whereby the contrast with the largest ratio-to-scale is dropped after each stage, and the procedure applied to the remaining contrasts (Hamada & Balakrishnan 1998: 7).

2.2.5 Juan & Pena (1992)

Juan and Pena proposed a different estimator for the contrast standard error, τ, calculated as follows: Let MAD_0 be the median of the *k* unsigned contrasts, and then iteratively calculate the median of the unsigned contrasts which do not exceed $wMAD_0$ for some constant $w > 2$. The process is stopped once the median stops changing, and this median is then called $IMAD_0$ (iterated median absolute deviation). An estimator for τ is then given by

$$\hat{\tau}_{\text{IMAD}} = \frac{\text{IMAD}_0}{a_w}$$

where a_w is a correction factor, given in their *Table 1* for a range of *w*. Juan and Pena recommended using $w = 3.5$ and $a_w = 0.6578$ (Hamada & Balakrishnan 1998: 11).

In their comparative study, Hamada and Balakrishnan concluded Juan and Pena's procedure to perform better for situations where a large number of active effects are present (1998: 22).

2.3 Non-sequential directed methods

2.3.1 Daniel (1959)(1)

One of the earliest methods to identify active effects (and still the standard method in practice) was proposed by Daniel in 1959. It is a graphical procedure which avoids the need for estimating the process standard error σ (Hamada & Balakrishnan 1998: 4).

The procedure consists of plotting the unsigned contrasts on half-normal probability paper and then drawing a straight line through the bulk of the small contrasts. Contrasts falling on the line are then deemed inert, while those falling of the line are flagged as active contrasts (Hamada & Balakrishnan 1998: 1).

This, however, implies assuming a sparsity of active effects, which may not hold true in a specific situation. Another objection which can be raised against Daniel's method is its subjectivity: there is no formal rule to indicate what signifies a point (contrast) falling "off the line" (Hamada & Balakrishnan 1998: 1). It led to this method sometimes half-condescendingly being referred to as the "fat pencil" technique.

2.3.2 Lenth (1989)

Assuming a sparsity of effects, Lenth (1989) proposed a robust estimator for the contrast standard error τ, which he named the *pseudo standard error (PSE)*. Letting $c_1, c_2, \ldots, c_m$ be the estimates for the m contrasts of interest, he calculates a preliminary estimate of τ as

$$s_0 = 1.5 \times \underset{j}{\text{median}} \left| c_j \right|$$

and the *PSE* of the contrasts is then defined as

$$PSE = 1.5 \times \underset{|c_j|<2.5s_0}{\text{median}} |c_j|$$

The *PSE* is therefore a trimmed median including only the contrasts within 2.5 times the preliminary estimate of τ, and "is a fairly good estimate of τ when the effects are sparse" (Lenth 1989: 470), even though it will still be an overestimation (Lenth 1989: 472). The square of *PSE* has an approximate chi-squared distribution (SAS Institute 2003).

To test which of the contrasts (if any) are active, a *margin of error (ME)* can be calculated for the c_i:

$$ME = t_{0.975;\, d} \times PSE$$

where $t_{0.975;\, d}$ is the 97.5th percentile from a t distribution with d degrees of freedom. Lenth suggests a value of $d = m/3$ and computes the percentiles from the t distribution using the fractional degrees of freedom (without rounding).

With the *ME*, 95% confidence intervals can be constructed for the contrasts, using $c_i \pm ME$. This however implies making a number of inferences simultaneously, increasing the probability of a Type I error and possibly leading to erroneous conclusions. For this reason, Lenth also defined a *simultaneous margin of error (SME)*:

$$SME = t_{\gamma;\, d} \times PSE$$

where

$$\gamma = \left(1 + 0.95^{1/m}\right)/2$$

This gives an exact (not conservative) *SME*, as the contrast estimates are independent of each other (Lenth 1989: 470).

Lenth further suggests displaying the contrasts graphically, and adding reference lines at $\pm ME$ and $\pm SME$, rather than performing formal hypothesis tests or constructing confidence intervals. Any contrasts of which the bars extend beyond the *SME* reference lines are then deemed as active, and those within the *ME* lines as inactive (Lenth 1989: 470).

In a simulation study done by Hamada and Balakrishnan, they concluded Lenth's method to perform better for cases where there are a large number of active effects (Hamada & Balakrishnan 1998: 22).

2.3.3 Modified Lenth (1989)

Hamada and Balakrishnan also suggested a modified version of Lenth's method. Instead of using a trimmed median to calculate the *PSE* as an estimate for τ, one can proceed sequentially "by dropping the largest contrast and applying the procedures on the remaining contrasts whose sample size is one less" (Hamada & Balakrishnan 1998: 13).

As with Lenth (1989), they found this modified version to also perform better for a large number of active effects, but providing little improvement over the original method except when the number of active effects becomes large (Hamada & Balakrishnan 1998: 22).

2.3.4 Berk & Picard (1991)

Hamada & Balakrishnan found the method proposed by Berk & Picard (1991) as one of the better directed methods and concluded it to be "one of the best methods" overall (1998: 22). Berk & Picard's method is ANOVA-based, similar to proposals made by Taguchi, Voss and Quinlan (Berk & Picard 1991: 79), and consists of pooling the smallest effect sums of squares to calculate a baseline "pseudo-error term", and then performing the usual *F*-tests on the sums of squares of the larger effect estimates.

Specifically, the ratio of the larger effect sums of squares to the mean square of the 60% smallest effect sums of squares is calculated

$$\frac{M_{(l)}}{\sum_{i=1}^{m} M_{(i)}}$$

where $M_{(m)}$ is the 60th percentile of the ordered effect mean squares. These ratios are then compared to the *F* critical values given in *Table 1* of the article by Berk & Picard (1991: 81), not the usual *F* percentiles. They stress the point that their pseudo-error term should not be interpreted as a variance estimate, and that "apparent 'error' estimates should be used with great caution, if at all" (Berk & Picard 1991: 88).

The pooling of the smallest effects involves the assumption of effect sparsity, i.e. an implicit assumption that only a minority of the effects can be active and that the smaller effects are null. This assumption may not necessarily hold true, in which case it may lead to the reduction of power and provide problems with the interpretation of the observed effects (Berk & Picard 1991: 80). However, "[b]ased on a sensitivity study, [they do] feel that reserving roughly 60% of the total effects for construction of the baseline [pseudo-error term] leads to an excellent procedure" (Berk & Picard 1991: 80). They further noted that, as their method is suited for screening experiments where the sparsity of effects assumption is typically relied on, the results should in any case be followed up with further experimentation (Berk & Picard 1991: 83).

2.3.5 Dong (1993)

Dong proposed a method which is very closely related to that of Lenth (1989), also calculating a *pseudo standard error (PSE)* and *simultaneous margin of error (SME)*, but "[basing] it on the trimmed mean of squared contrasts rather than the trimmed median of the unsigned contrasts" (Hamada & Balakrishnan 1998: 12).

Dong's estimator for the contrast standard error τ is given by

$$s_{DONG} = \sqrt{\frac{1}{m}\sum_{\{|c_j|<2.5s_0\}} c_j^2}$$

where s_0 is as defined in Lenth (1989) and m is the number of squared contrasts included in the summation. To test the individual contrasts for being active, he then standardizes the contrasts using

$$\frac{|c|_{(i)}}{s_{DONG}}$$

where $|c|_{(i)}$ is the ith order statistic of the unsigned contrasts, and compares it to the critical value $t_{\gamma,m}$ with γ being defined as in Lenth (1989). Dong showed that s_{DONG} has a smaller *mean squared error (MSE)* than Lenth's *PSE* (Hamada & Balakrishnan 1998: 12), and is therefore an improved estimator of τ.

In their comparative study, Hamada & Balakrishnan (1998: 22) found Dong's procedure to perform better for situations where only a small proportion of the effects are active.

2.3.6 Modified Dong (1993)

Dong also suggested "calculating s_{DONG} [iteratively] until it stops changing when there are a large number of active effects" (Hamada & Balakrishnan 1998: 12). Hamada & Balakrishnan included this modified version in their study and established that it, as with the original version, performs better for a small number of active effects. They concluded that it "has almost the same power as [Dong (1993)] for small [number of active effects], and actually performs worse as [the number of active effects] increases. Thus, there is no real benefit offered by the iteration in estimating τ." (Hamada & Balakrishnan 1998: 22)

2.3.7 Schneider, Kasperski & Weissfeld (1993)

In the past, one of the more popular types of techniques to estimate the error variance had been to successively pool the smallest unsigned contrasts as an estimate for σ. An major drawback of such methods, however, is that it "systematically underestimates the error variance" (Schneider et al. 1993: 19). Schneider, Kasperski and Weissfeld (1993) then proposed a technique to reduce the bias in the estimate when successive pooling is used to estimate the error variance.

Their technique consists of pooling the *n* smallest unsigned contrasts and then treating it as a Type II right censored sample. They find an estimate for the error variance by using the maximum likelihood estimate from a Type II right censored normally distributed sample, as follows:

Letting *H* denote the hazard function

$$H(x) = \frac{\phi(x)}{1-\Phi(x)}$$

where $\phi(x)$ is the standard normal density function and $\Phi(x)$ the standard normal distribution function. They then compute

$$H(M/\sigma) = \frac{\phi(M/\sigma)}{1-\Phi(M/\sigma)}$$

where *M* is the absolute value of the n^{th} smallest contrast effect (i.e. $M = |c_n|$), and s_n (the standard deviation of the *n* smallest effects) is used as an preliminary estimate for σ.

Letting *m* be the total number of contrasts, and $r = m - n$, they proceed by computing

$$A(\sigma) = \frac{r}{n} H(M/\sigma) M$$

and then obtain an estimate of σ by using iteratively the equation

$$\sigma = \frac{A(\sigma)}{2} + \sqrt{\frac{A(\sigma)^2}{4} + s_n^2}$$

The value of σ is then used in the next iteration until the change in the value of σ in subsequent iterations becomes small, at which point the process will be terminated (Schneider et al. 1993: 26).

Once an estimate for σ have been obtained, one can compute the modified *t* statistics

$$t_i = \frac{c_i}{\hat{\sigma}}$$

To determine which contrasts are active, they construct the limits

$$T = c_i \pm \tau\hat{\sigma}$$

in which τ is defined as

$$\tau = \tau_{\alpha/2} = \sqrt{\frac{z_{\alpha/2}^2}{1 - \frac{\gamma}{m} z_{\alpha/2}^2}}$$

where

$$\gamma = \left[2\left(\frac{n}{m}\right) - u\phi(u) + u^2\left(\frac{m\phi(u)}{r} - u\right)\right]^{-1}$$

in which $u = \Phi^{-1}(n/m)$. The τ value is then used in *T* to determine the inner and outer limits ($\pm\tau_{0.025}$ and $\pm\tau_{\alpha/2}$ for the inner and outer limits, respectively, where $\alpha = 1-(1-0.05)^{1/m}$).

The estimated error variance (and the *t* statistics based thereon) from this method are less biased than that obtained from other successive pooling methods, and "yields results similar to those obtained using normal probability and Bayes plots" (Schneider et al. 1993: 19). Both the method and its results are quite robust to the choice of *n*, but Schneider et al. recommends choosing *n* as approximately equal to half the total number of contrasts (Schneider et al. 1993: 25).

In the comparative simulation study done by Hamada and Balakrishnan (1998: 22), they found the method proposed by Schneider, Kasperski & Weissfeld to perform better in situations where there are only a small number of active effects. They concluded it to be, together with Daniel's (1959) method, one of the better directed methods suited for a small number of active effects.

2.3.8 Modified Schneider, Kasperski & Weissfeld (1993)

Hamada & Balakrishnan (1998) suggested a modification on Schneider, Kasperski & Weissfeld (1993): instead of choosing *n* equal to about half the total number of effects and then treating the smallest *n* contrasts as a Type II censored sample, the procedure can be done by sequentially dropping the largest contrast and performing the calculations on the remaining contrasts. However, the critical values will then need to be recalculated for the reduced sample size at each stage (Hamada & Balakrishnan 1998: 13).

In their comparative study, they found this modified version to perform better for a larger number of active effects, although it provides little if any improvement over the original version except for such cases where there are a large number of active effects (Hamada & Balakrishnan 1998: 22).

2.3.9 Venter & Steel (1996)

Venter & Steel (1996) proposed a procedure which uses a hypothesis testing approach to identify active effects. They test the overall hypothesis

$$H_0 : c_1 = c_2 = ... = c_n = 0$$

against the alternative that at least one of the effects is non-zero (i.e. active). If this overall null hypothesis is rejected, their procedure allows the identification of the effects which are significantly different from zero (Venter & Steel 1996: 161).

To test the overall null hypothesis, Venter & Steel uses the ratios

$$V_i = \frac{|c|_{(i+1)}}{g_i(|c|_{(1)}, \ldots, |c|_{(i)})}, \qquad i = 1, \ldots, n-1$$

where $g_i(|c|_1, \ldots, |c|_i)$ is an scale-equivariant function of its arguments, for which Venter & Steel proposes the form

$$g_i(|c|_1, \ldots, |c|_i) = \left(\frac{1}{i} \sum_{j=1}^{i} |c|_j^2 \right)^{\frac{1}{2}}$$

A *p*-value is calculated for each of the successive V_i ratios by simulation: *B* independent random vectors $\left(T_1^{(b)}, \ldots, T_n^{(b)}\right)$, $b = 1, \ldots, B$ are generated, for which all T_i are independently and identically distributed values from a standard normal distribution. The order statistics $(|T|_{(1)}^{(b)}, \ldots, |T|_{(n)}^{(b)})$ and corresponding ratios $(V_1^{(b)}, \ldots, V_n^{(b)})$ are calculated for each of the *B* random vectors. The *p*-value for the *i*th *V* ratio can then be approximated by

$$\hat{P}_i = \frac{1}{B} \sum_{b=1}^{B} I\left(V_i < V_i^{(b)}\right)$$

where $I(A)$ is an indicator function for which the value is one if event *A* is true, and zero otherwise.

Venter & Steel assumes effect sparsity, i.e. that at least the *l* smallest contrasts correspond to inactive effects, and therefore confine their test to the $n - l - 1$ largest contrasts. They define

$$S_l = \min\{P_i : l \le i \le n-1\}$$

and then calculate the distribution of S_l under H_0 by using simulation in a similar way as for calculating the *p*-values. They denote the distribution of S_l under the null hypothesis by s_l, and use its quantiles as critical values for the hypothesis test. If $S_l \leq s_l(\alpha)$ (where $s_l(\alpha)$ is the αth quantile of s_l), H_0 is rejected at a significance level of α (Venter & Steel 1996: 162).

If the overall hypothesis test is rejected, the active effects are "identified by the first index $\hat{q} \geq l$ such that $P_{\hat{q}} \leq s_l(\alpha)$" (Hamada & Balakrishnan 1998: 13).

2.4 Sequential composite methods

2.4.1 Bissell (1989)

Bissell proposed a method which uses Bartlett's test for the homogeneity of variance to identify possible active effects. It uses the statistic

$$B = \ln\left(\left(\frac{1}{k}\right)\sum M_i\right) - \left(\frac{1}{k}\right)\sum \ln(M_i)$$

where k is the number of contrasts, and M_i is the ith mean square. It is used sequentially, and compared to critical values which are based on an appropriate *F* distribution. The critical value at the ith stage is based on the assumption of the remaining $k - i + 1$ being inactive (Hamada & Balakrishnan 1998: 9).

It is therefore a composite method in the sense that it indicates whether there are active effects present or not, without explicitly identifying which of the effects are active (if any).

2.4.2 Bissell (1992)

Bissell also suggested another composite method for identifying active effects: the procedure employs the statistic

$$S_M \big/ \bar{M}$$

where S_M and $\bar{M}$ are the standard deviation and mean of the k smallest mean squares (the M_i), respectively. Under the null hypothesis of no active effects, all the mean squares have the same scaled χ^2 distribution. It then follows that

$$\left(\frac{k-1}{2}\right)\left(S_M \big/ \bar{M}\right)^2$$

is distributed approximately $\chi^2_{(k-1)}$, for which the critical values is given in *Table 12* of Bissell's article. He suggested sequentially leaving out a number of the mean squares for which the effects are obviously active, and then retesting the remaining effects using the critical values for the reduced sample size.

However, Hamada and Balakrishnan notes the possible criticism "that the χ^2 approximation does not account for the fact that the estimate $\bar{M}$ is used rather than the true mean" (1998: 10).

In their empirical study they also found that Bissell's method "appears promising ... for up to four active effects but then its performance seriously degrades for six active effects ... [because] the variance of the mean squares will tend to decrease when there are too many active effects (i.e., the roles of the inert and active contrasts are switched) while their mean increases resulting in small values" (Hamada & Balakrishnan 1998: 22). Bissell's procedure therefore seems to be more suited to situations where there are only a few active effects.

2.5 Non-sequential composite methods

2.5.1 Probability Plot Correlation Coefficient

In their comparative study, Hamada & Balakrishnan (1998) also considered a method proposed by Filliben in 1975. It builds on Daniel's idea of constructing a probability plot of the effect contrasts, and consists of measuring the linearity of such a plot by calculating the correlation coefficient between the ordered contrasts and the median standard normal order statistics:

$$R_{CORR} = \frac{\sum_{i=1}^{k}\left(\tilde{z}_{(i)k} - \bar{\tilde{z}}\right)\left(c_{(i)} - \bar{c}\right)}{\sqrt{\sum_{i=1}^{k}\left(\tilde{z}_{(i)k} - \bar{\tilde{z}}\right)^2}\sqrt{\sum_{i=1}^{k}\left(c_{(i)} - \bar{c}\right)^2}}$$

where $\tilde{z}_{(i)k}$ is the median standard normal order statistics, and $c_{(i)}$ is the ordered contrasts. However, the method which Hamada & Balakrishnan actually tested in their simulation study, was a half-normal version which uses the unsigned contrasts and expected standard half-normal order statistics (denoted by $|c_i|$ and $|z|_{(i)k}$, respectively). The correlation coefficient then becomes

$$R_{CORR} = \frac{\sum_{i=1}^{k}\left(|z|_{(i)k} - |\bar{z}|\right)\left(|c_i| - |\bar{c}|\right)}{\sqrt{\sum_{i=1}^{k}\left(|z|_{(i)k} - |\bar{z}|\right)^2}\sqrt{\sum_{i=1}^{k}\left(|c_i| - |\bar{c}|\right)^2}}$$

Small values of R_{CORR} would point to a lack of linearity in the half-normal probability plot, which would indicate that at least some of the effects may be active. They also note that the procedure can be performed sequentially, leaving out the largest contrast at each stage and recalculating R_{CORR} and the critical values for the reduced sample of contrasts.

2.5.2 Half-Normal Shapiro-Wilk Test

In their article, Hamada and Balakrishnan (1998: 14) proposes a half-normal version of the Shapiro-Wilk test to detect active effects in a 2^k factorial design. They work with the unsigned contrasts, and defined the *Best Linear Unbiased Estimator (BLUE)* of τ based on the *m* smallest unsigned contrasts as

$$\hat{\tau}_{\text{BLUE}} = \frac{\underset{\sim}{\mu}^T \Sigma^{-1} \left|\underset{\sim}{c}\right|_{()}}{\underset{\sim}{\mu}^T \Sigma^{-1} \underset{\sim}{\mu}}$$

where $\left|\underset{\sim}{c}\right|_{()}$ is the vector of *m* smallest contrasts, $\underset{\sim}{\mu}$ is the vector of the means of the *m* smallest standard half-normal order statistics in a sample of size *k*, and Σ is the variance-covariance matrix of these order statistics.

Furthermore, the *MLE* of τ based on the unsigned contrasts is given by

$$\hat{\tau}_{\text{MLE}} = \sqrt{\frac{1}{k}\sum_{i=1}^{k} |c_i|^2}$$

and therefore their Shapiro-Wilk type goodness-of-fit test is given by

$$\text{HSW} = \frac{\hat{\tau}_{\text{BLUE}}}{\hat{\tau}_{\text{MLE}}}$$

Small values of HSW would suggest the presence of active effects. Hamada and Balakrishnan suggest using HSW sequentially by removing the largest unsigned contrast at each stage, and recalculating HSW for the reduced sample size. Critical values for HSW are based on empirical analysis (Hamada & Balakrishnan 1998: 15).

2.5.3 Dispersion Test

Hamada & Balakrishnan (1998: 15) also introduces a *dispersion test* based on the distribution of the effect contrasts under the null hypothesis of no active effects. If the null hypothesis holds true, the unsigned contrasts $|c_i|$ should have a half-normal

distribution, and their order statistics $|c|_{(i)}$ should on average be close to $\tau|z|_{(i)k}$ (where τ and $|z|_{(i)k}$ are the contrast standard error and the expected value of the *i*th standard half-normal order statistic out of *k*, respectively).

Taking Lenth's *pseudo standard error (PSE)* as a robust estimator of τ, they use the *m* smallest contrasts to calculate the statistic

$$D_m = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{|c|_{(i)}}{PSE\,|z|_{(i)k}} - 1\right)^2$$

An active effect would cause the $|c|_{(i)}/PSE$ to deviate significantly from its expected value $|z|_{(i)k}$, leading to large values of D_m, in which case the null hypothesis of no active effects must be rejected. "This statistic can also be used sequentially with critical values being calculated for the reduced sample size at each stage" (Hamada & Balakrishnan 1998: 15).

2.6 Hybrid methods

2.6.1 Benski (1989)

Benski suggested a procedure which Hamada and Balakrishnan in their article calls a "*hybrid method*", in the sense that it consists of two parts: Firstly, a "global" composite test to find out if there are any active effects present. If the test indicates the presence of active effects, one proceeds to the second part, in which the active effects are identified.

The first part of Benski's method is a modified Shapiro-Wilk test for normality, using the ratio of the squared estimated slope of the probability plot regression line to the standard deviation of the contrasts, giving the statistic

$$W' = \frac{\left(\sum_{i=1}^{k} z_{(i)k} c_{(i)} \right)^2}{\left(\sum_{i=1}^{k} z_{(i)k}^2 \sum_{i=1}^{k} \left(c_{(i)} - \bar{c} \right)^2 \right)}$$

where $\bar{c}$ is the average of the ordered contrasts and the $z_{(i)k}$ are the expected standard normal order statistics in a sample of size *k*. Small values of W' would indicate that the effect contrasts do not all have the same (zero) mean, and that at least one of the effects are active.

Once the presence of active effects had been established, Benski employs an outlier test to identify the specific effects which are active. Letting d_F be the *interquartile range* of the contrasts, he proposes the interval

$$\left(-2d_F \; ; \; +2d_F \right)$$

as an acceptable spread for the inert effects. This interval assumes a zero mean for the inert effects, and all contrasts falling outside the interval are identified as possibly being active.

Benski then proposed the following method to test the possibly active effects formally: "If the Shapiro-Wilk test is rejected, combine the *p*-values of both tests and declare the largest contrast active if the combined test is rejected. Then, drop the largest contrast and perform the same procedure on the remaining contrasts" until no more active effects are found (Hamada & Balakrishnan 1998: 8).

2.6.2 Loh (1992)

Another hybrid method was proposed by Loh, with the motivation to formally extend the graphical normal probability plot procedure. Similar to Benski's test, the first part

of Loh's procedure is designed to determine if there are any active effects present. If so, one proceeds to the second part in which the specific active effects are identified.

In the first part of the procedure, a least-squares regression line is fitted to the normal probability plot of all the contrasts, and compared to the regression line through the inert contrasts (those of which the absolute size are less than twice the interquartile range d_F). This test is rejected for large ratio values, and the outliers (effects of which the magnitude is more than $2d_F$) are then considered to be potentially active.

Loh then uses the Scheffé prediction interval based on the regression line fitted through the inert effects identified in the first part of the test: the outliers falling outside this prediction interval are then identified as being active (Hamada & Balakrishnan 1998: 11-12).

2.6.3 Modified Loh (1992)

Hamada and Balakrishnan notes "that some computation is required in finding the set of contrasts used in the normal plot" when using Loh's method (1998: 12). They suggest working with the unsigned contrasts, which leads to a half-normal version of the procedure.

For the half-normal version, the inliers (inert effects) are taken to be those not greater than four times the median of the unsigned contrasts. Let $\hat{\beta}_1$ be the slope of the regression line fitted through the origin to all the unsigned contrasts in the half-normal probability plot, and $\hat{\beta}_2$ the slope of the regression line fitted through the origin to the unsigned contrasts corresponding to the inliers. The first part of the test then consists of evaluating the statistic

$$R = \frac{\hat{\beta}_1}{\hat{\beta}_2}$$

which will be rejected for large values of *R*. If rejected, one proceeds to identify the active effects causing the rejection as those outliers exceeding the prediction interval based on the fitted line to the inliers (from the first part of the test). Letting S_2 be the

root mean squared error of the fitted line through the inliers, outliers corresponding to active effects will have

$$\left| |c|_{(l)} - \hat{\beta}_2 |z|_{(l)k} \right| > S_2 \left(k' F_{k',m-1;\gamma} \right)^{\frac{1}{2}} (1+w)^{\frac{1}{2}}$$

where

$$k' = \text{ceiling}\left[\frac{k}{4} \right]$$

and

$$w = \frac{|z|^2_{(l)k}}{\sum_{i=1}^{m} |z|^2_{(i)k}}$$

and $|c|_{(l)}$ and $|z|_{(l)k}$ are the absolute value of the lth outlying contrast and its expected standard half-normal order statistic, respectively (Hamada & Balakrishnan 1998: 13).

2.7 Sparsity of effects assumption

The majority of the procedures discussed in this chapter depend (some heavily) on the assumption that there will be a *sparsity of effects* in the situations where they will be employed. In practice, this assumption may or may not be realistic (Hamada & Balakrishnan 1998: 29) – if not, it may severely affect the research conclusions based on the results obtained from these methods.

Generally, researchers have found that the sparsity of effects assumption holds up in practice, with only about 20% of the total number of effects being active (Hamada & Balakrishnan 1998: 2). However, for a specific situation this assumption may still prove to be wrong. There is a need for methods which are robust to a violation of the assumption, even just as backup procedures, if nothing else. The next chapter will be exploring some variable selection techniques (which do not depend on sparsity of effects) with application to the identification of active effects in unreplicated 2^k factorial experiments.

Chapter 3:
New Proposals

3.1 Introduction

The identification of active effects by way of a factorial experiment can be viewed as a variable selection problem. Much work had been done on the topic of variable selection, but these developments have normally bypassed any application to the unreplicated 2^k factorial problem.

This chapter therefore firstly takes a brief look at the reasons why traditional variable selection procedures fail to provide an acceptable solution in the 2^k factorial environment. Thereafter follows some new proposals: using the *Lasso* procedure as a variable selection technique, and its implementation by means of *Least Angle Regression (LARS)*, and the *Logworth method*.

3.2 Traditional variable selection procedures

Application of the conventional methods of variable selection usually involves the fitting of a number of models (often of different sizes) to the data, and then selecting the "best" one by way of one or more selection criteria. Many of these selection measures are based on the error sum of squares of the model under consideration,

usually penalizing in some way for p (number of model parameters) to allow direct comparison of models of different sizes.

One of the popular model selection measures is Mallow's C_p. Letting p be the number of variables included in the model under consideration, and n the number of observations on which it was estimated. The C_p value for the model is then given by

$$C_p = \frac{SSE_p}{\hat{\sigma}^2_{Full}} + 2p - n$$
$$= \left(1 - R^2\right)\left(\frac{SST}{\hat{\sigma}^2_{Full}}\right) + 2p - n$$

where R^2 is the coefficient of determination, and $\hat{\sigma}^2_{Full}$ the error variance of the full model containing all the independent variables. Clearly C_p cannot be used in the unreplicated 2^k factorial setting where the full model provides a perfect fit and consequently an error variance of zero.
Two other commonly used measures are *Akaike's Information Criterion (AIC)*, which, for a model with n observations and p variables is given by

$$AIC = n\ln\left(\frac{SSE}{n}\right) + 2p$$
$$= n\ln\left(1 - R^2\right) + n\ln\left(\frac{SST}{n}\right) + 2p$$

and *Schwarz's Bayesian Information Criterion (SBC)*

$$SBC = n\ln\left(\frac{SSE}{n}\right) + p\ln(n)$$
$$= n\ln\left(1 - R^2\right) + n\ln\left(\frac{SST}{n}\right) + p\ln(n)$$

The forms of *AIC* and *SBC* are very similar, differing only in the definition of the penalty term (for the number of variables included in the model). Model selection is done by choosing the model with the minimum *AIC* and/or *SBC*. For regression models based on unreplicated 2^k factorial designs, it will work up to the point where the full model is fitted. With $p = n$, R^2 will be one, and both *AIC* and *SBC* will break down.

There is another bigger problem, however. For such small designs, R^2 tends to one at a much faster rate than the growth of the penalty term ($2p$ and *AIC* and $p\ln(n)$ for *SBC*). In other words, the increase in the penalty term fails to counter the effect of the increase in the model fit. Consequently, both *AIC* and *SBC* will be monotone decreasing, never reaching a minimum turning point.

		effects						
response		**A**	**B**	**C**	**AB**	**AC**	**BC**	**ABC**
67	*(1)*	-1	-1	-1	1	1	1	-1
79	*a*	1	-1	-1	-1	-1	1	1
61	*b*	-1	1	-1	-1	1	-1	1
75	*ab*	1	1	-1	1	-1	-1	-1
59	*c*	-1	-1	1	1	-1	-1	1
90	*ac*	1	-1	1	-1	1	-1	-1
52	*bc*	-1	1	1	-1	-1	1	-1
87	*abc*	1	1	1	1	1	1	1

Source: Vining 1998: 374

Table 3.1: Data from an unreplicated 2^3 factorial (steel springs with cracks experiment).

This phenomenon can be illustrated using Vining's steel spring data (1998: 374), given in *Table 3.1*. Adjusting the response for its mean (thus eliminating the need to estimate the intercept), a forward selection technique can be used to produce regression models of sizes $p = 1, 2, \ldots, 7$. The *AIC* and *SBC* values for each of these models are given in *Table 3.2*.

p	**AIC**	**SBC**
1	54.54	54.70
2	44.76	44.99
3	32.08	32.40
4	28.94	29.34
5	12.52	13.00
6	-515.43	-514.87
7	-523.18	-522.54

Table 3.2: AIC *and* SBC *values for models with* p *variables.*

Both *AIC* and *SBC* are monotone decreasing, failing to reach a minimum turning point even for a mean-centred model. Clearly these measures cannot be used for variable selection when working with unreplicated 2^k factorial designs, for it will always select the full model (and consequently declare all effects active).

3.3 The Lasso

The *Lasso (Least Absolute Shrinkage and Selection Operator)* is a linear model estimation method proposed by Tibshirani (1996). It is similar to ridge regression in that it minimizes the error sum of squares subject to a constraint on the estimated coefficients. More specifically, whereas ridge regression employs a quadratic constraint, the lasso puts an upper limit on the sum of the absolute values of the coefficients.

Tibshirani considers the Lasso to be superior to *ordinary least squares (OLS)* regression for two reasons: Firstly, an overspecified OLS model often has little bias but large variance, adversely affecting its prediction accuracy. This can be improved by shrinking or setting to zero (i.e. dropping from the model) some of the coefficients, trading some bias for a lower model variance. Secondly, OLS models may sometimes have a large number of small coefficients, adding little value to the model and complicating the interpretation of the effects. The Lasso addresses these problems by shrinking some coefficients and setting others to zero (Tibshirani 1996: 267). This brings about the additional benefit of being able to use it as a variable selection technique, by including only the nonzero lasso effects in the final model.

To formally define the method, let $\left(x^i, y_i\right)$, $i = 1, 2, \ldots, N$ be a set of data as for the usual regression set-up, with $x^i = \left(x_{i1}, \ldots, x_{ip}\right)'$ the predictor variables and y_i the response for the ith observation. Further, assume that the x_{ij} are standardized so that

$$\sum_i \frac{x_{ij}}{N} = 0 \quad \text{and} \quad \sum_i \frac{x_{ij}^2}{N} = 1$$

Now, if the y_i are centred so that $\bar{y} = 0$, the Lasso estimate $\hat{\beta}$ is defined to be the set of $\hat{\beta} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)'$ which minimizes

$$\sum_{i=1}^{N}\left(y_i - \sum_j \beta_j x_{ij}\right)^2 \quad \text{subject to} \quad \sum_j \left|\beta_j\right| \le t$$

where $t \geq 0$ is the tuning parameter controlling the amount of shrinkage applied to the coefficient estimates. If $\hat{\beta}_j^0$ are the full OLS estimates and $t_0 = \sum \left| \hat{\beta}_j^0 \right|$, values of $t < t_0$ will shrink the estimates towards zero, causing variables with coefficients equal to zero to, in effect, be dropped from the model (Tibshirani 1996: 268).

Tibshirani suggests three different methods for estimating the bound *t*: *K*-fold cross validation, generalized cross validation, and a third method based on Stein's unbiased estimate of risk. Only the first method will be discussed here, as it is the only one implemented in the *R* package (***lars***) at the time of writing, and it is the method used for the execution of the Lasso in the simulation study (chapter 4).

Suppose that

$$Y = \eta(X) + \varepsilon$$

where $E(\varepsilon) = 0$ and $\text{var}(\varepsilon) = \sigma^2$. Then the *mean-squared error (ME)* of the estimate $\hat{\eta}(X)$ is defined as

$$ME = E\left\{\hat{\eta}(X) - \eta(X)\right\}^2$$

and the *prediction error (PE)* as

$$PE = E\left\{Y - \hat{\eta}(X)\right\}^2 = ME + \sigma^2$$

This prediction error is then estimated using *K*-fold cross validation: letting

$$s = \frac{t}{\sum \hat{\beta}_j^0}$$

be the normalized parameter, the prediction error is estimated over a grid of *s*-values (0 to 1 inclusive), with the value of $\hat{s}$ yielding the lowest PE being selected. For the simulation study (chapter 4), a value of $K = n$ was used (which corresponds to *leave-one-out* cross validation) to estimate the absolute bound *t*.

Finding the Lasso solutions is a quadratic programming problem, which can be done with standard numerical analysis algorithms (Efron et al. 2004: p3?). Another way, however, is to use *least angle regression (LARS)*: with a single modification, LARS gives the entire path of lasso solutions for all values (from 0 to infinity) of the bound *t*. LARS will therefore be discussed in detail in the following section.

3.4 Least Angle Regression (LARS)

Least angle regression (LARS) is a relatively new model selection algorithm, much of its appeal lying in the fact that it is a less greedy version than the traditional forward selection methods. It is also computationally efficient, requiring "only the same order of magnitude of computational effort as Ordinary Least Squares applied to the full set of covariates" (Efron et al. 2004: p1?). For the complete set of LARS steps with $m < n$ covariates, $O\left(m^3 + nm^2\right)$ computations are required (Efron et al. 2004: p34?).

As noted in the previous section, the Lasso can be implemented by a single modification to the basic LARS algorithm, decreasing the amount of computation required to find the Lasso solutions by an order of magnitude (compared to previous methods). Using *m* to denote the number of covariates, at most *m* steps are required to produce the full set of Lasso solutions (Efron et al. 2004: p4-5?). At each of the successive steps, the covariate(s) with the biggest current correlation is added to the model.

Assuming for the regression data $\left(x^i, y_i\right)$, $i = 1,\ 2,\ \ldots,\ N$ that

$$\sum_{i=1}^{n} y_i = 0, \quad \sum_{i=1}^{n} x_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} x_{ij}^2 = 1 \quad \text{for} \quad j = 1,2,\ldots,m$$

and that the covariate vectors $x_1,\ x_2,\ \ldots,\ x_m$ are linearly independent, the algorithm begins at $\hat{\mu}_0 = 0$, where $\hat{\mu}_0$ denotes the vector of effect means (i.e. all variable coefficients start off at zero). The vector of current correlations

$$\hat{c} = X'\left(y - \hat{\mu}_A\right)$$

is then computed, after which the greatest absolute correlation $\hat{C}$ and the active set A (which is the set of indices corresponding the current greatest absolute correlations) are found:

$$\hat{C} = \max_j \left\{ \left| \hat{c}_j \right| \right\} \quad \text{and} \quad A = \left\{ j : \left| \hat{c}_j \right| = \hat{C} \right\}$$

Letting

$$s_j = \text{sign}\left\{ \hat{c}_j \right\} \quad \text{for} \quad j \in A$$

we compute

$$X_A = \left(\ldots \; s_j x_j \; \ldots \right)_{j \in A}$$

and also

$$G_A = X_A' X_A \quad \text{and} \quad A_A = \left(1_A' G_A^{-1} 1_A \right)^{-\frac{1}{2}}$$

where 1_A denotes a vector of 1's of length $|A|$ (the size of A).

The *equiangular vector* is then found:

$$u_A = X_A w_A \qquad \text{where} \qquad w_A = A_A G_A^{-1} 1_A$$

and used to compute the inner product vector

$$a \equiv X' u_A$$

Then, for the next step of computations, the mean effect vector $\hat{\mu}_A$ is updated to

$$\hat{\mu}_{A+} = \hat{\mu}_A + \hat{\gamma} u_A$$

with $\hat{\gamma}$ being defined as

$$\hat{\gamma} = \min^{+}_{j \in A^c} \left\{ \frac{\hat{C} - \hat{c}_j}{A_A - a_j}, \frac{\hat{C} + \hat{c}_j}{A_A + a_j} \right\}$$

where $\min^{+}$ indicates that the minimum is taken only over the positive components for each choice of *j*. The procedure is terminated when all covariates are included in the model (Efron et al. 2004: pp5-7?).

3.4.1 For the orthogonal design case

For the unreplicated 2^k factorial design with coded factor levels (± 1) as discussed in chapter 1, the mathematics of the LARS algorithm can be simplified somewhat. With $\hat{c}$, $\hat{C}$, A, and X_A defined as before, the matrix G now becomes (Efron et al. 2004: p18?)

$$G = X_A' X_A = I_A$$

because the columns of X are orthogonal, each of length 1. Then

$$A_A = \left(1_A' G^{-1} 1_A\right)^{-1/2} = \frac{1}{\sqrt{|A|}}$$

where $|A|$ denotes the length of A (i.e. the number of covariates in the model) as before, and

$$w_A = A_A G^{-1} 1_A = \frac{1}{\sqrt{|A|}} 1_A = \begin{bmatrix} \frac{1}{\sqrt{|A|}} \\ \frac{1}{\sqrt{|A|}} \\ \vdots \end{bmatrix}$$

Therefore the equiangular vector can be simplified to

$$u_A = X_A w_A = \frac{1}{\sqrt{|A|}} X_A 1_A$$

and the inner product becomes

$$a = X' u_A = \begin{bmatrix} X_A' u_A \\ X_A^{c'} u_A \end{bmatrix} = \begin{bmatrix} \frac{1}{|A|} 1_A \\ 0 \end{bmatrix}$$

Further, the parameter $\hat{\gamma}$ can be simplified to

$$\hat{\gamma} = \min_{j \in A^c}^{+} \left\{ \frac{\hat{C} - \hat{c}_j}{A_A - a_j}, \frac{\hat{C} + \hat{c}_j}{A_A + a_j} \right\}$$

$$= \min_{j \in A^c}^{+} \left\{ \frac{\hat{C} - \hat{c}_j}{\frac{1}{\sqrt{|A|}} - 0}, \frac{\hat{C} + \hat{c}_j}{\frac{1}{\sqrt{|A|}} + 0} \right\}$$

$$= \min_{j \in A^c}^{+} \left\{ \frac{\hat{C} - \hat{c}_j}{\frac{1}{\sqrt{|A|}}}, \frac{\hat{C} + \hat{c}_j}{\frac{1}{\sqrt{|A|}}} \right\}$$

and then, as before, $\hat{\mu}_A$ is updated to $\hat{\mu}_{A+} = \hat{\mu}_A + \hat{\gamma} u_A$ for each successive step of computations.

For the Lasso modification, the algorithm is as described above but for the exception that, when occasionally the coefficient of a covariate previously entered into the model is shrunken to zero, it should be dropped from the model (without excluding it from entering the model again at a later stage). Computationally, this will cost at most $O\left(m^2\right)$ operations per variable dropped (Efron et al. 2004: p34?).

3.5 The Logworth method

The term *logworth* originates from the field of data mining, and is simply a way of rescaling p-values to facilitate easier comparison by human analysts. P-values (from fitted models, coefficients, etc.) are often very small: 10^{-16} and 10^{-20} may be deemed by a human data analyst to be approximately equal in size because both are relatively close to zero (on the p-value range [0 ; 1]), but the first value is in fact 10 000 times larger than the second!

The logworth measure transforms a p-value to the negative of its natural logarithm, i.e.

$$-\ln(\text{p-value})$$

which, for the example above, would rescale the two p-values to 36.84 and 46.05, respectively. Smaller p-values translate to bigger logworth values, and by comparing them on the logworth scale, a human analyst can then clearly see that there is a difference between them.

A more important advantage, however, is that the direct comparison of p-values are not limited by the underlying distributions they originate from, nor by issues of differing degrees of freedom, unlike, for example, F-ratios from different sized regression models. While p-values do depend on the specific distribution from which they are derived (and its parameters, degrees of freedom, etc.), they are expressed on a scale which is independent of their origins.

For the method as set forth here, the logworth measure is used to select the optimal model from the set of best models of each size: For the data set $\left(x^i, y_i\right)$, $i = 1, 2, \ldots, N$ under consideration, with each x^i a vector of p covariates, all subsets of regression models of size $1, 2, \ldots, p-1$ are fitted and evaluated according to the p-value of the F-ratio

$$F = \frac{MSR}{MSE}$$

where MSR and MSE denotes the mean sum of squares for regression and mean sum of squares for error, respectively. The set of $p-1$ models comprising the best

model (i.e. the one with the minimum p-value) of each subset size are then found and compared by their logworth scores. For such a set the logworth reaches a maximum at the point where adding more variables starts worsening the model fit. The model size corresponding to such a peak is then taken to be indicative of the number of active effects present.

The logworth scores of a set of $p-1$ models may have more than one such peak, however. To deal with this problem, and also with the possibility of no active effects, threshold logworth scores can be implemented. Analogous to the *F-to-enter* of traditional forward selection methods, one can calculate the threshold score

$$-\ln(\alpha)$$

where α is the probability of making a Type I error. Any local maximum in the logworth scores smaller than the threshold score may then be rejected as spurious. Further, for reasons of model parsimony, one may wish to select the model corresponding to the first local maximum in the logworth scores (the approach followed by this author).

Graphs of the logworth scores can be plotted for a visual comparison, with boundary lines drawn at $-\ln(\alpha)$. The Logworth method, including these graphs, was implemented in the *R* function *logworth.exact* (see Appendix B). Using the data given in *Table 3.1*, the output from *logworth.exact* is as follows:

```
$activenum
[1] 3

$model
  (Intercept)             A             B            AC
-9.322424e-17  1.150000e+01 -2.500000e+00  5.000000e+00

$min.log.pvals
[1]  5.956703  7.743787  9.238200  7.447069  6.960632 36.736801
```

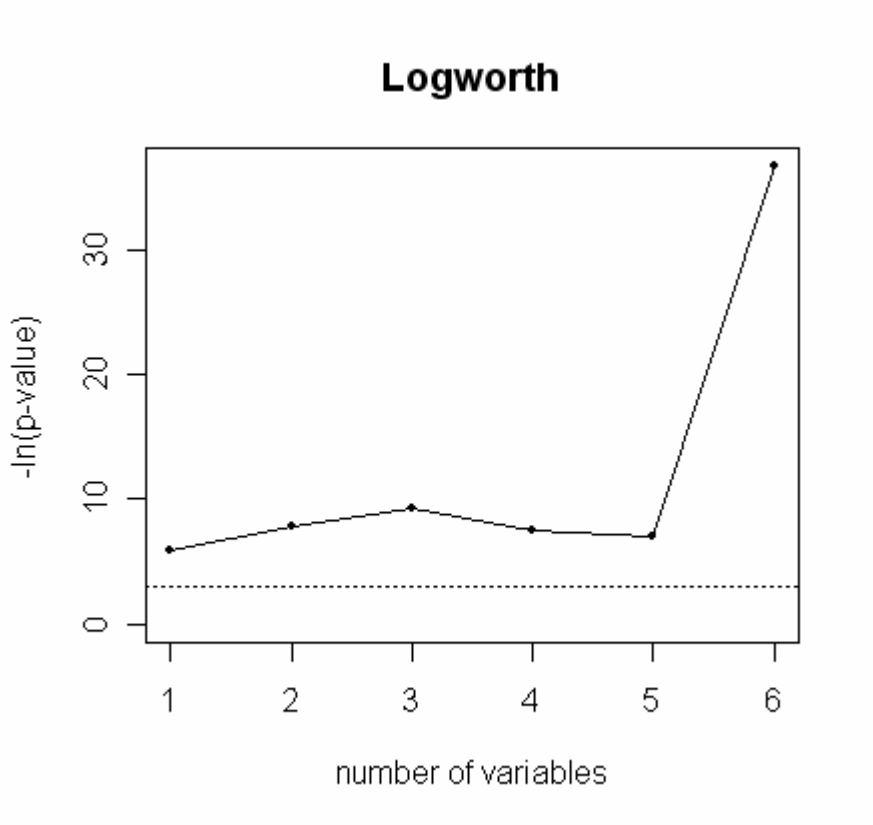

Figure 3.1: Graph of the logworth scores for Vining's unreplicated 2^3 factorial experiment. The broken line indicates the boundary at –ln(0.05). A local peak is reached for the model with three effects.

One disadvantage of the Logworth method, as set out above, is that it is computationally very taxing. Fitting all possible subsets of models to a data set involves a large number of calculations, and increases exponentially for an increasing number of covariates: even for a small 2^4 factorial design (with three main effects, three second-order interactions, and a single third-order interaction), a total of 32 767 linear regression models have to be fitted. For a 2^5 design, the number of possible subset models increases to 2 147 483 647!

Thus, for the purpose of a big simulation study, using the aforementioned approach proved to be prohibitively expensive with regards to computer processing time. Another approach had to be used: the orthogonal structure of the design matrix implies independence of the effects. (This assumption was also made for the data simulation process which will be discussed in chapter 4). Consequently, the specific effects included in each subset model will be independent, and the best model of size q ($q = 1, 2, \ldots, p-1$) will simply be the one containing the q covariates most

correlated with the response variable. Therefore, the best model of size *q* can be found by merely including the first *q* variables from the Lasso sequence of models, thereby decreasing the number of regression models to be fitted to $p-1$.

The equivalence of this second approach can again be illustrated using the 2^3 design in *Table 3.1*. The output generated by the R routine *logworth* (Appendix B) was as follows:

```
$activenum
[1] 3

$min.log.pvals
[1] 5.956703 7.743787 9.238200 7.447069 6.960632      Inf
```

The logworth scores are exactly equal to those from the exact approach, and once again the first local maximum as found at $q=3$, indicating three active effects.

For the simulation study in chapter 4, this shortcut approach had been used. The threshold score have been set at $-\ln(0.05)$, which in retrospect maybe was not stringent enough. However, it placed the Logworth method on a more “equal footing” with the other procedures, where a probability of 0.05 for committing a Type I error had been used wherever explicitly specified.

The following chapter reports on the performance of these newly proposed methods, compared to some of the more established procedures discussed in chapter 2.

Chapter 4:
Simulation Study

4.1 Introduction

The different methods for detecting active effects in unreplicated 2^k factorial experiments was put to the test in a comparative simulation study. The results from the study are reported here, after which a discussion thereof follows in chapter 5.

Firstly, some practical considerations are discussed: how the data was simulated, selection of methods for inclusion in the study, and error rate measures used to compare the different procedures. An overview of the results from the different scenarios is then given – the interested reader may refer to *Appendix C* for the complete set of tables and figures mentioned below.

4.2 Simulated data

An unreplicated 2^4 factorial structure was used throughout as the basic construct for the simulation study. Including all main effects and interactions, this involved the estimation of fifteen factor effects in total.

For each simulation run, response data was simulated by sampling from a multivariate normal distribution with a mean vector as specified by the user, and a 15x15 identity matrix as the covariance matrix. Thus the assumed paradigm is one of homogeneity of variance and no correlation between the effects (both active and inactive).

All computions were performed in the *R* environment, by using routines written entirely by the author, unless otherwise indicated. These routines can be found in *Appendix B* of this assignment. All graphs were produced in *R*, and tables prepared in *Excel*.

4.3 Selection of methods

The classic procedures to analyze unreplicated factorial experiments have been thoroughly investigated by a number of authors in the past, most notably by Hamada & Balakrishnan (1998). It was decided to take the conclusions drawn by these authors as a basis and proceed with further study, rather than to simply replicate their work.

Therefore the choice of methods included in this simulation study reflect the experience and findings of past investigators, as well as the interest of this author and the new possible solutions proposed in chapter 3. Giving much weight to the work done by Hamada & Balakrishnan (1998), it was decided to include the procedures as proposed by Lenth (1989), Dong, Berk & Picard (1991), Schneider, Kasperski & Weissfeld (1993), and Venter & Steel (1996). All of these methods are also included in the ADX (Analysis of Designed Experiments) interface of the SAS system, which can be seen as an indication of their dependability and effectiveness in practice.

As no previous study have been done on the Lasso as a variable selection method in the unreplicated 2^k factorial design setup, it was of interest to see how it would perform compared to the more established procedures. The Lasso deals gracefully even with situations where $p >> n$, including only the $n-1$ effects most correlated with the response variable, which appeared to be a promising aspect for coping with unreplicated 2^k factorial designs where p is nearly equal to n.

The newly proposed Logworth method was also included in the simulation study to investigate its performance characteristics when compared on an equal basis to the well-known techniques. If it outperforms the other procedures, even just in certain scenarios, it may prove to be a helpful and user-friendly method for future experimenters.

4.4 Error rates

For a comparative measurement of the performance characteristics of the methods, two different error rates, as defined in the article by Hamada & Balakrishnan (1998: 17), will be used. Suppose there are k active effects, and, letting p_i denote the proportion of simulated samples for which i ($i = 0, 1, \ldots, n-1$) effects have been identified as active by a specific procedure, the *experimentwise error rate* (EER) will be defined as

$$EER = 1 - p_k$$

i.e. the proportion of simulated samples for which the wrong number of effects have been identified as active. The method with the lowest EER will therefore have been the one with the highest rate of correctly identifying the true number of active effects.

Secondly, define the *individual error rate* (IER) as the average proportion of inactive effects declared active (Hamada & Balakrishnan 1998: 17), i.e.

$$\sum_{i=k+1}^{n-1} p_i \left(\frac{i}{n-k-1} \right)$$

The IER will therefore be an indicator of how greedy a specific procedure is: a higher IER corresponds to a higher tendency to identify inert effects as active.

4.5 Different scenarios investigated

The comparative simulation study was rather extensive, investigating a number of possible scenarios for an unreplicated 2^4 factorial experiment. Taking a cue from the work done by Hamada & Balakrishnan (1998), it was decided to include the following broadly grouped spectrum of situations:

1) no active effects
2) small number of active effects
3) large number of active effects
4) small number of large active effects

21 different scenarios are included in the four categories mentioned above. These will be explained in more detail in the sections which follows. For the scenarios with active effects present, all active effects were of the same sign and magnitude. According to Hamada & Balakrishnan, "[a]ctive effects with the same magnitude ... provide bounds on the performance of when the effects have different magnitudes" (1998: 19).

For each of the scenarios, 10 000 response data samples were simulated and evaluated with each of the seven chosen procedures, in order to provide a stable and accurate assessment of their performance characteristics.

4.5.1 No active effects

Firstly, the off-the-shelf performance of the seven methods was investigated by using a scenario where there are no active effects. Data was simulated as from an unreplicated 2^4 factorial experiment for a zero mean effect vector. The results are summarized in *Table 4.1*.

Lenth, Dong, and Schneider emerged as the top three methods for this scenario, correctly identifying zero active effects in at least 92% of all the simulated samples. The method proposed by Venter and the Lasso lags somewhat behind. All these procedures therefore have a relatively low EER.

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.979	0.920	0.552	0.953	0.845	0.712	0.002
	1	0.013	0.058	0.259	0.038	0.012	0.000	0.035
	2	0.004	0.014	0.116	0.007	0.014	0.001	0.038
	3	0.002	0.006	0.053	0.002	0.015	0.004	0.058
	4	0.001	0.002	0.016	0.001	0.014	0.006	0.082
	5	0.001	0.000	0.005	0.000	0.014	0.010	0.113
	6	0.000	0.000	0.000	0.000	0.017	0.013	0.138
	7	0.000	0.000	0.000	0.000	0.016	0.019	0.148
	8	0.000	0.000	0.000	0.000	0.017	0.019	0.145
	9	0.000	0.000	0.000	0.000	0.018	0.022	0.117
	10	0.000	0.000	0.000	0.000	0.018	0.021	0.074
	11	0.000	0.000	0.000	0.000	0.000	0.026	0.036
	12	0.000	0.000	0.000	0.000	0.000	0.033	0.011
	13	0.000	0.000	0.000	0.000	0.000	0.049	0.001
	14	0.000	0.000	0.000	0.000	0.000	0.067	0.001
	EER	0.021	0.080	0.448	0.047	0.155	0.288	0.998
	IER	0.002	0.008	0.049	0.004	0.060	0.207	0.437

Table 4.1: Experimental and individual error rates of the seven different methods when there are no active effects (off-the-shelf performance). Cells contain the proportions effects declared active by the specific method for 10 000 simulated data sets, using an unreplicated 2^4 factorial design (16 runs).

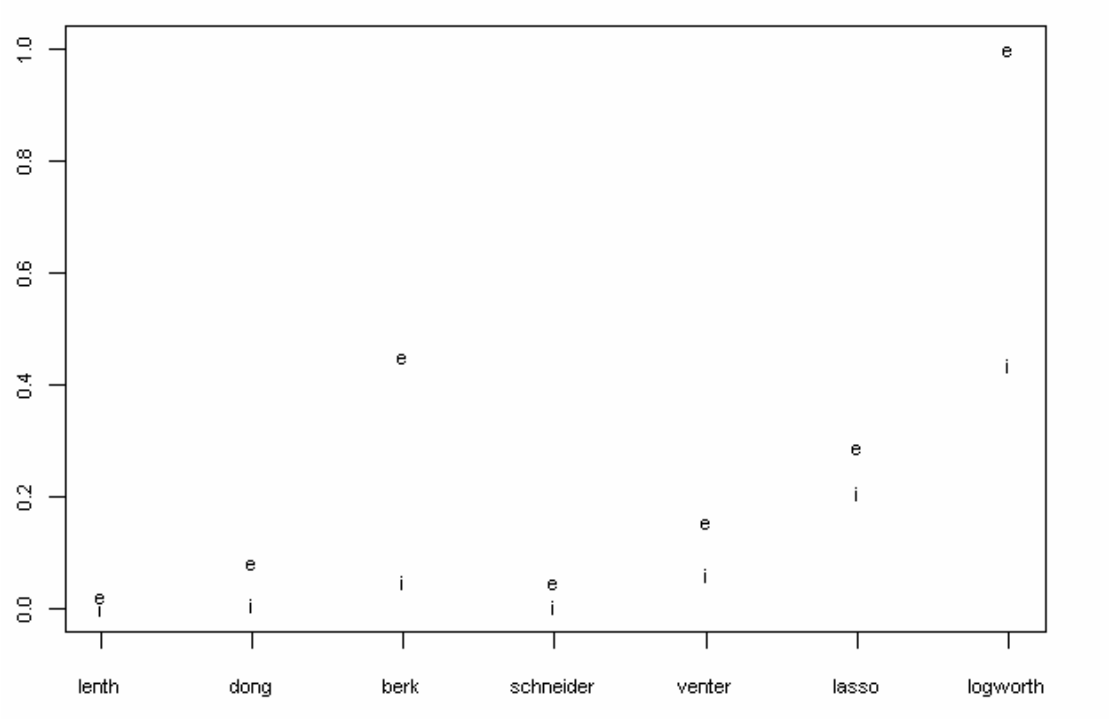

Figure 4.1: Experimental and individual error rates (denoted by "e" and "i", respectively) of the seven different methods when there is no active effects (off-the-shelf performance).

Berk correctly declared zero active effects for about half of the simulated samples, while the Logworth method performed terribly, declaring zero effects for only 0.2% of the simulated samples. Consequently, its EER is very large at 0.998.

In terms of the IER, all the methods, with the exception of the Lasso and Logworth, has very low IER, showing them to be quite conservative. The IER rates of the Lasso and Logworth are 0.207 and 0.437, indicating a higher tendency to declare some inert effects active. For a visual comparison, both the EER and IER rates were plotted in *Figure 4.1*.

The poor performance of Berk and especially the Logworth need not be of much concern, however: in practice, a situation with zero active effects is highly unlikely, even more so for screening experiments where these 2^k factorial designs are typically used. In such early stages of experimentation, the experimenter usually would be more concerned about Type II errors (incorrectly declaring some active effects to be inert and dropping them from subsequent experimentation) than Type I errors (incorrectly declaring some inert effects active) (Hamada & Balakrishnan 1998: 34). Type I errors will normally be confirmed in later, more extensive stages of experimentation, where the inactive variables can be sifted out and dropped with more convincing evidence.

4.5.2 Small number of active effects

To test the methods for the case where there are only a small number of active effects present, simulation were done using scenarios with one and three active effects, respectively, each ranging from 1σ to 6σ in magnitude. For a single active effect, the EER and IER rates are represented graphically in *Figures 4.2.1* and *4.2.2*, respectively. (Full results are given in *Tables 4.2.1-4.2.4* in Appendix C). On these and the following graphs, the plotting characters (1, 2, 4, and 6 for *Figures 4.2.1* and *4.2.2*) indicate the active effect mean-shift in number of standard deviations.

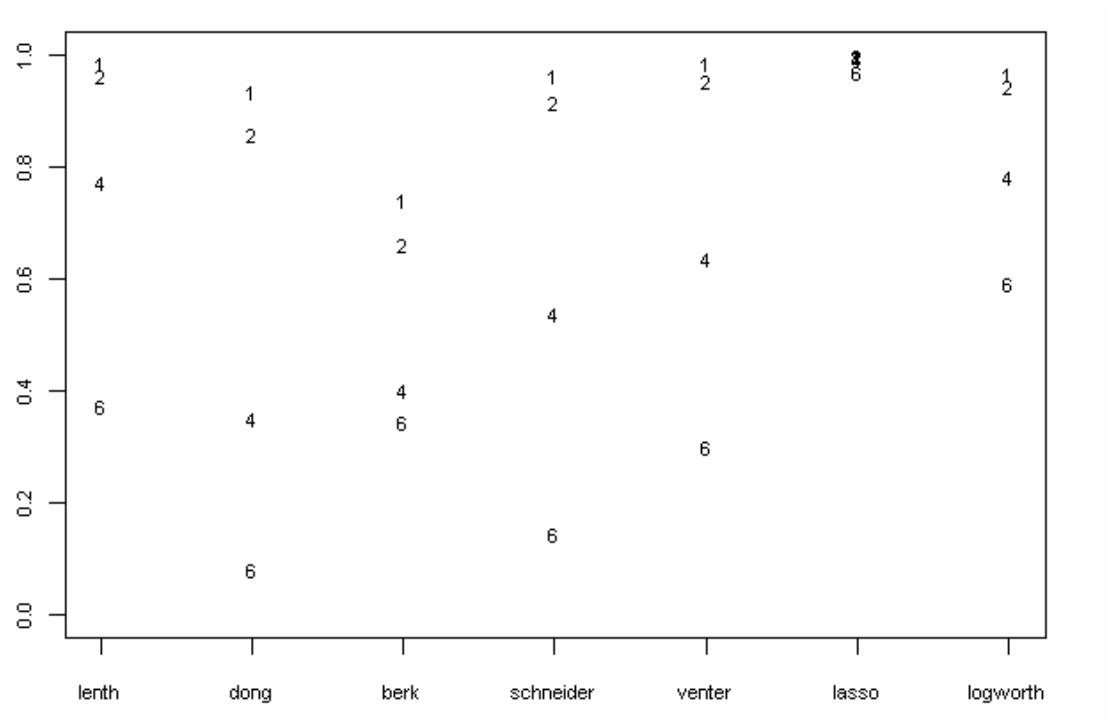

Figure 4.2.1: EER for situation with a single active effect present. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

The EER for a single very small (1σ mean-shift) effect are roughly the same for all the methods, with Berk performing slightly better than the rest. As the effect magnitude is increased, some differentiation begins to occur: overall, Berk performs the best, with Dong, Schneider and Venter catching up as the effect becomes larger. The Lasso performs the worst by far, with the Logworth only doing slightly better.

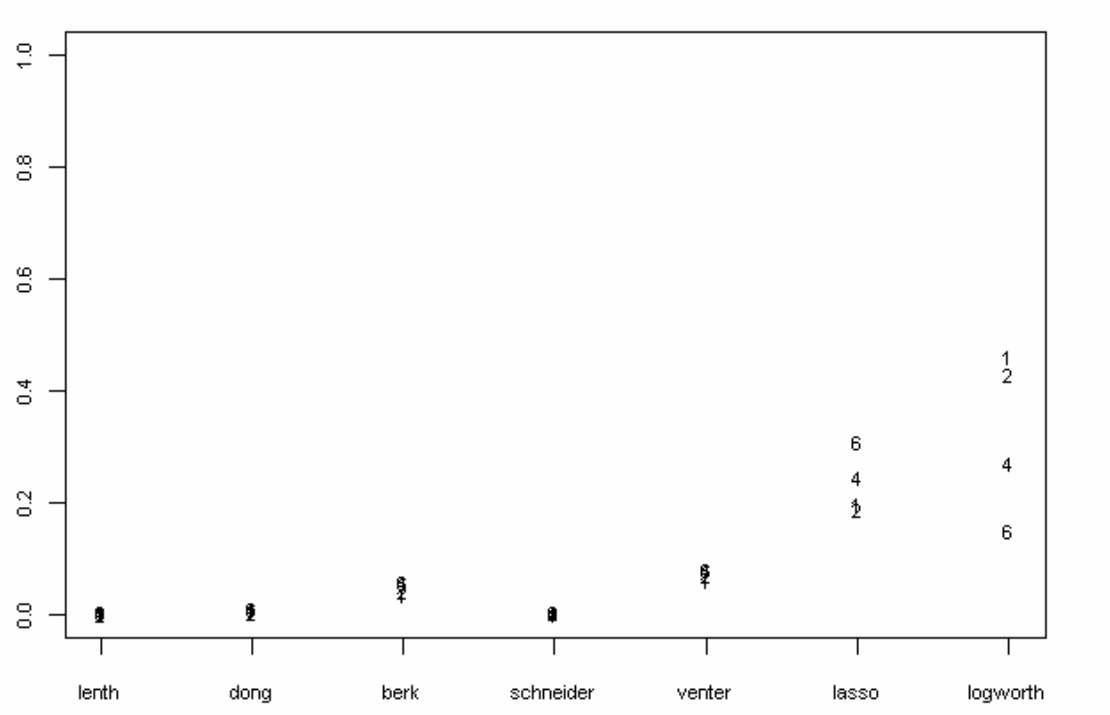

Figure 4.2.2: IER for situation with a single active effect present. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

From *Figure 4.2.2*, one can see the same trend as for the scenario with no active effects: once again the Lasso and Logworth appear to be the more greedy methods, having relatively large Type I error rates. The other methods perform about equally, having low IER rates for a single active effect of any size.

Changing the number of active effects to three does not make much of a difference. The EER and IER rates can be seen in *Figures 4.3.1* and *4.3.2*, with the full results contained in *Tables 4.3.1-4.3.4* (Appendix C).

Berk performs slightly worse for three effects of small magnitude (1σ, 2σ mean-shift) than for a single effect. Together with Dong and Venter, it is still one of the three best procedures in terms of EER. The Lasso and Logworth has the highest EER rates by far, as was the case for a single effect. Logworth outperforms the Lasso, though.

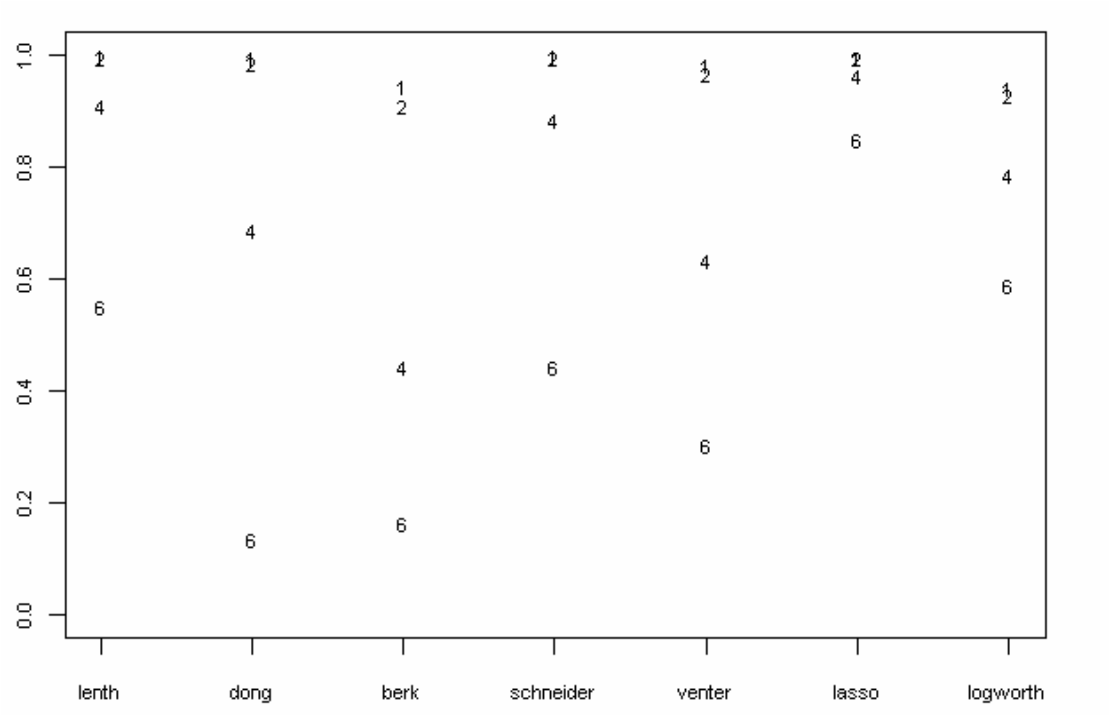

Figure 4.3.1: EER for situation with three equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

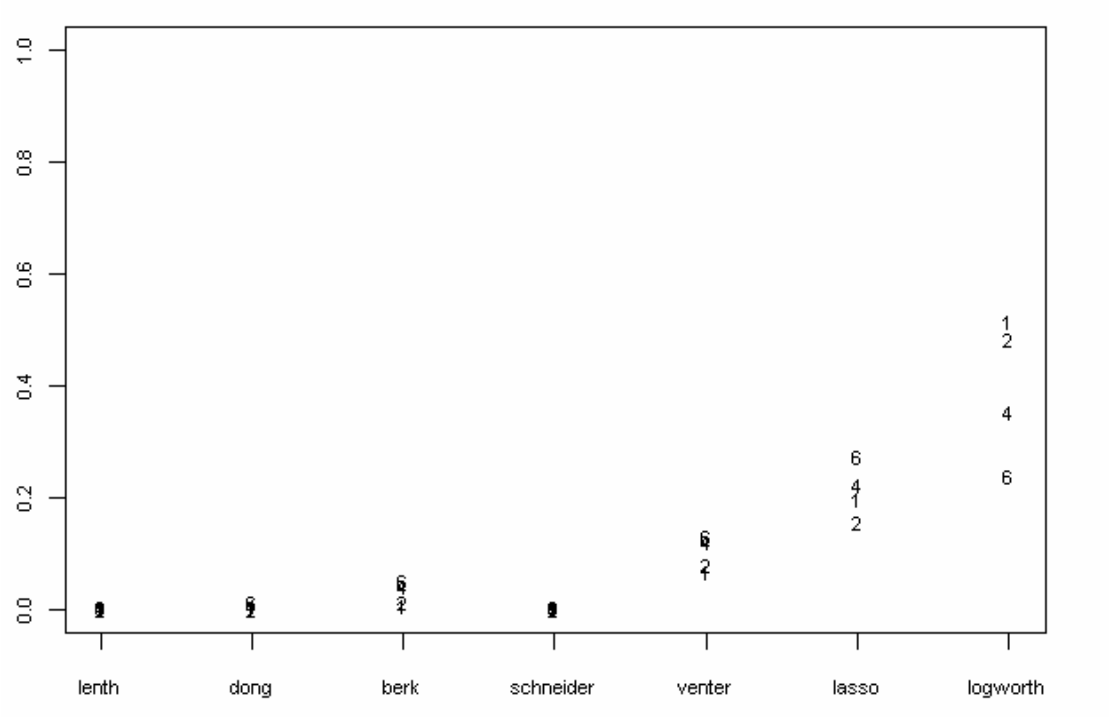

Figure 4.3.2: IER for situation with three equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

Figure 4.3.2 also continues the same trend seen in *Figure 4.2.2*: the Lasso and Logworth methods are still the more liberal ones, including on average more inert effects than the other methods. For three active effects, Venter becomes slightly less conservative than Lenth, Dong, Berk, and Schneider, but still has a relatively low IER.

4.5.3 Large number of active effects

A third possible situation is one where there are a large number of relatively small active effects present. To test the seven methods under such circumstances, simulations was done for detecting six and ten equally sized effects, ranging from 1σ to 6σ in magnitude. Results are given in *Figures 4.4.1-4.5.2* (full results contained in *Tables 4.4.1-4.5.4* in Appendix C).

The performance characteristics of the different procedures changed dramatically for the large number of effects scenario. As can be seen from *Figure 4.4.1*, the Logworth performs the best overall in terms of EER, especially for small effects (mean-shift of 1σ, 2σ). Schneider and the Lasso performs very poorly, while Berk and Venter only outperforms the Logworth when the effect mean-shift increases to 6σ.

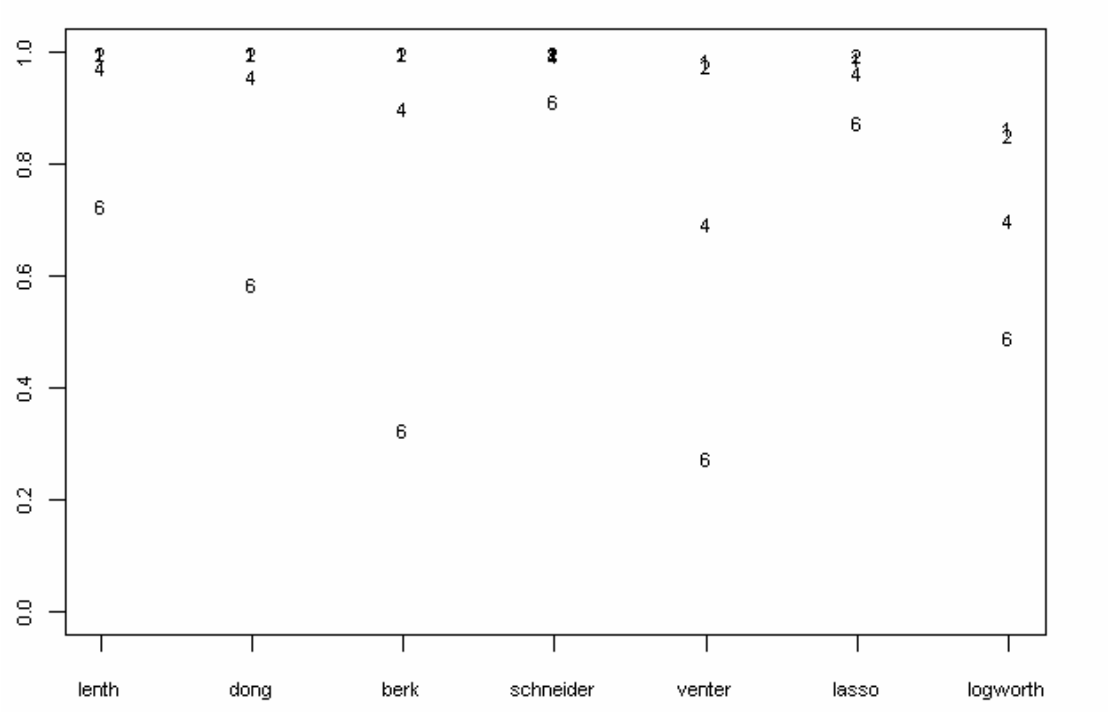

Figure 4.4.1: EER for situation with six equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

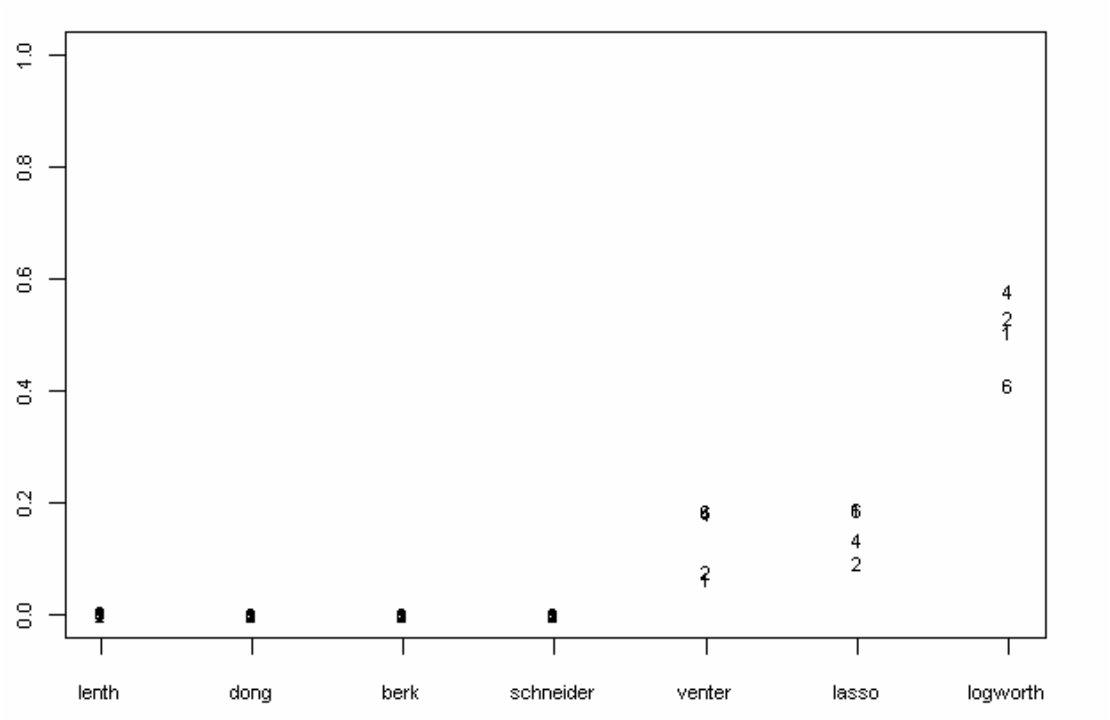

Figure 4.4.2: IER for situation with six equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

In terms of IER, the same general trend (as for a small number of active effects) is continued, with Lenth, Dong, Berk, and Schneider all maintaining IER rates close to zero, while Logworth are once again the least conservative method. Venter and the Lasso performs about equally – not overly conservative nor overly greedy.

Increasing the number of active effects to ten merely served to make these trends more distinct. The EER rates of Venter and the Logworth are similar (the Logworth outperforming Venter slightly) while the rest of the methods has EER rates bordering on or equal to one.

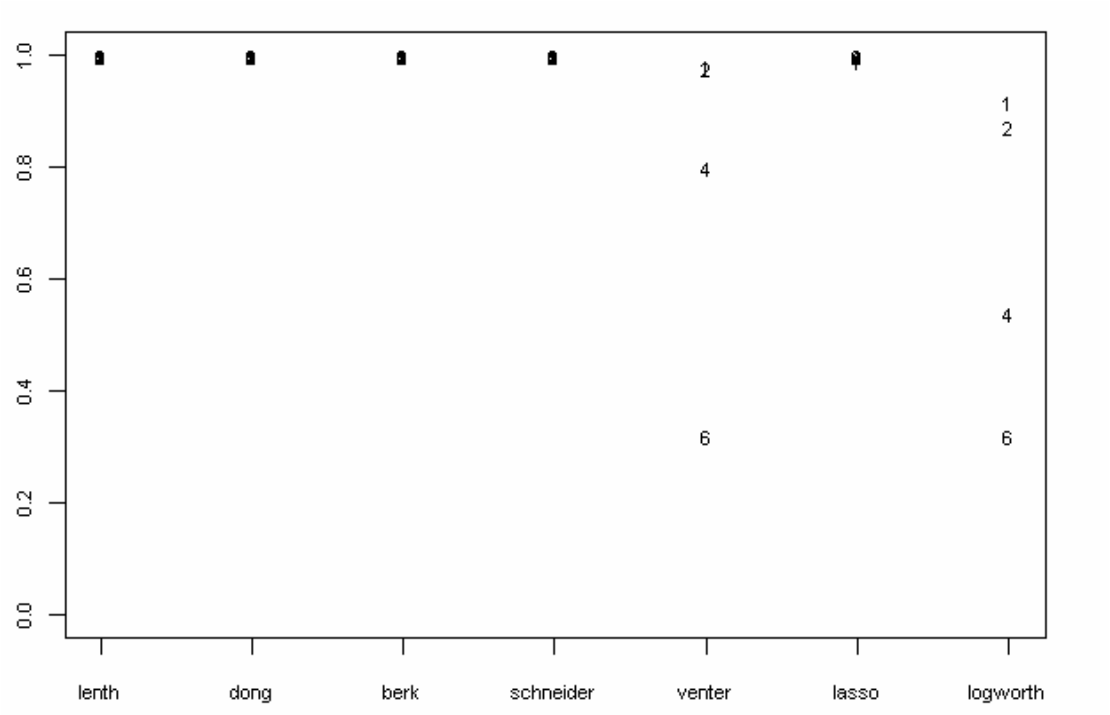

Figure 4.5.1: EER for situation with ten equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

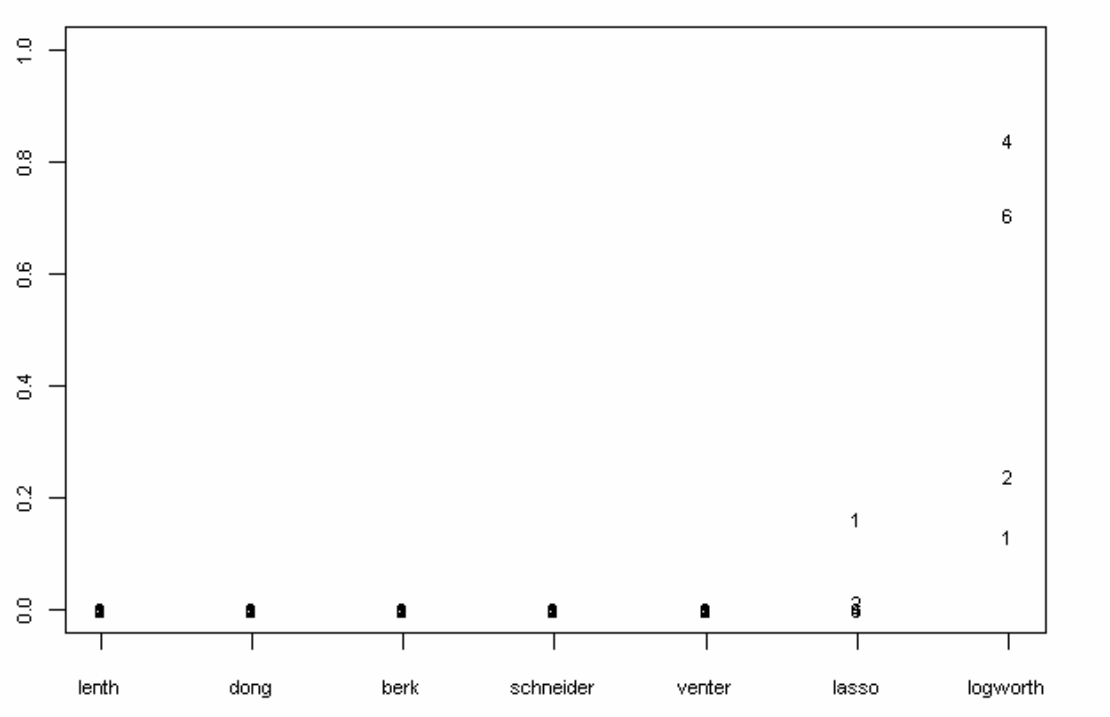

Figure 4.5.2: IER for situation with ten equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

For ten effects, the IER rates of the first five methods all equal zero, the effect size making no difference whatsoever. The only method with a distinctly higher IER is the Logworth, indicating that even for a large number of active effects it still fails to be conservative, declaring some inert effects active.

4.5.4 Small number of large effects

The last category of scenarios which were investigated was that of situations where there are only a few very large active effects. This will probably be the case for most experiments in practice, and therefore are of fundamental interest to be explored. Four different scenarios were simulated, two having a single active effect with mean-shift 10σ and 15σ, respectively, and two scenarios with three active effects (equally sized at 10σ and 15σ, respectively). Once again, 10 000 samples were generated per scenario.

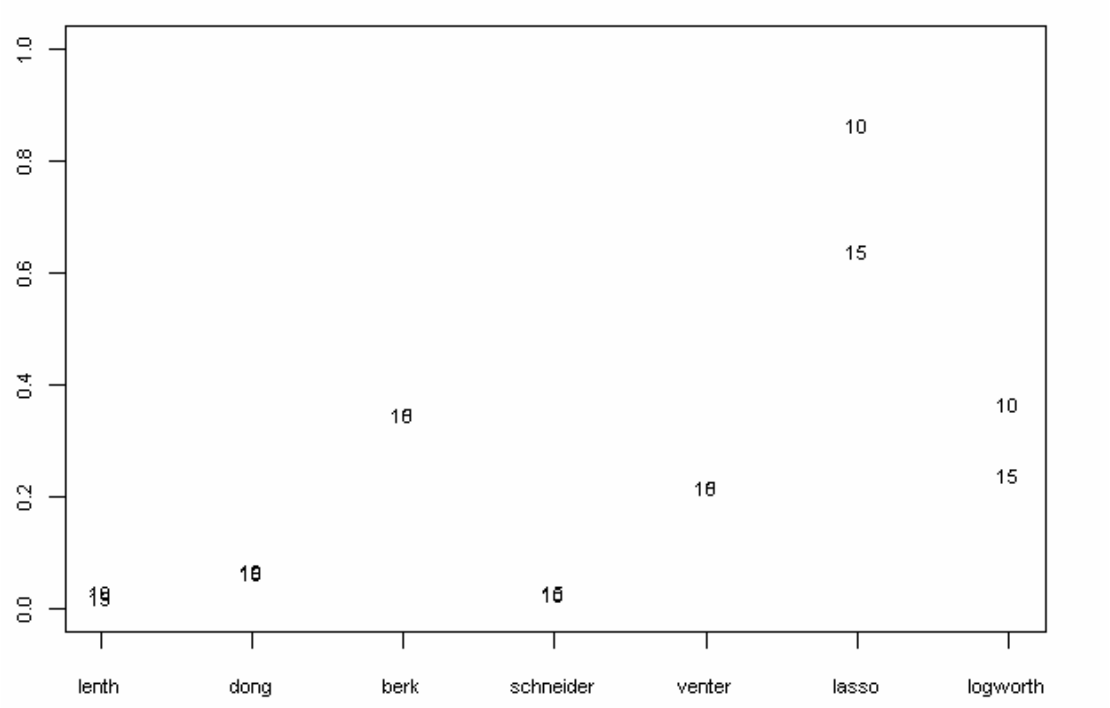

Figure 4.6.1: EER for situation with a single large active effect present. The numbers (10, 15) indicate the active effect mean-shift in number of standard deviations.

For a single large active effect, Lenth, Dong, and Schneider performed best in terms of their EER rates (*Figure 4.6.1* – for complete tables of results, refer to *Tables 4.6.1-*

4.6.2 in Appendix C). The Logworth performed quite competitively with Dong and Venter, showing a slight improvement in EER when the effect size was increased from 10σ to 15σ. Similarly, the Lasso also showed an improvement with the increase in effect size, but performed disappointingly compared to any other method.

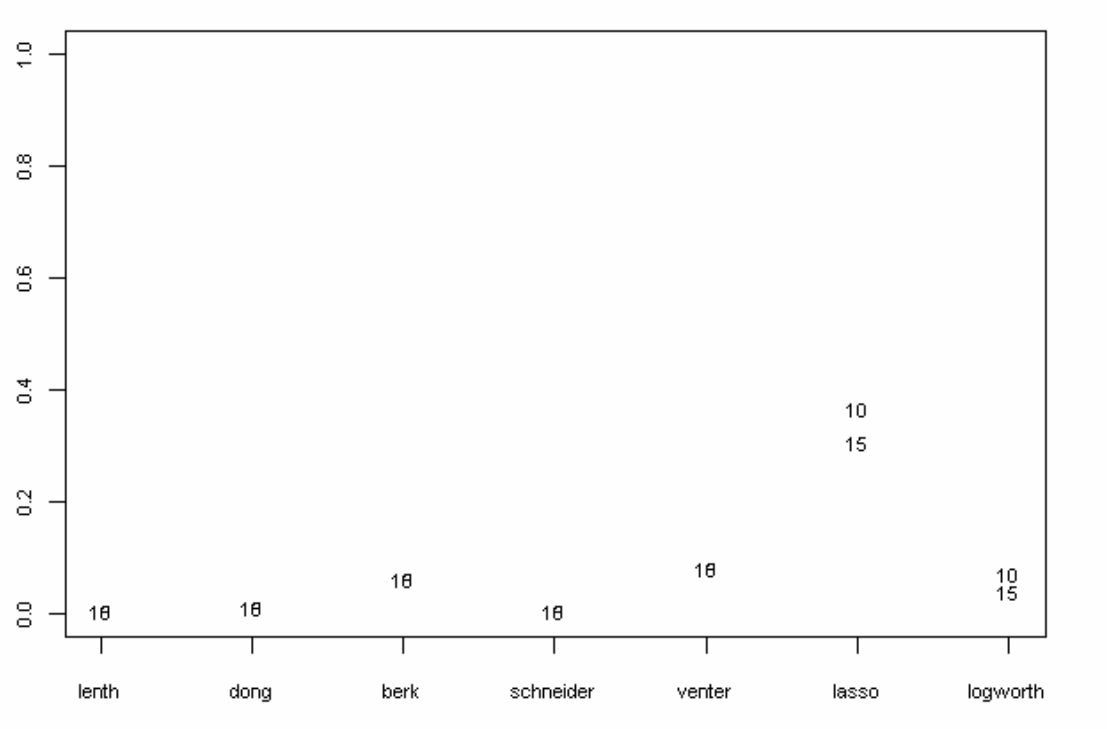

Figure 4.6.2: IER for situation with a single large active effect present. The numbers (10, 15) indicate the active effect mean-shift in number of standard deviations.

The IER rates for a single large effect is shown in *Figure 4.6.2*. In contrast with the previous scenarios, the Logworth has a very low IER, performing competitively with the first five methods and being only slightly worse than the usually more conservative procedures (Lenth, Dong, Schneider). The Lasso again performs poorly in terms of IER as well, showing itself to be the most liberal method for a large effect size.

When the number of active effects was increased to three, the same general trends were observed, but with a few notable exceptions (*Figures 4.7.1* and *4.7.2* – complete tables of results contained in *Tables 4.7.1* and *4.7.2* in Appendix C). Although the Lasso performed somewhat better, there was not much of an improvement in the EER rates of the Logworth and Venter. Berk performed slightly better, and Lenth, Dong, and Schneider consistently maintained low EER rates.

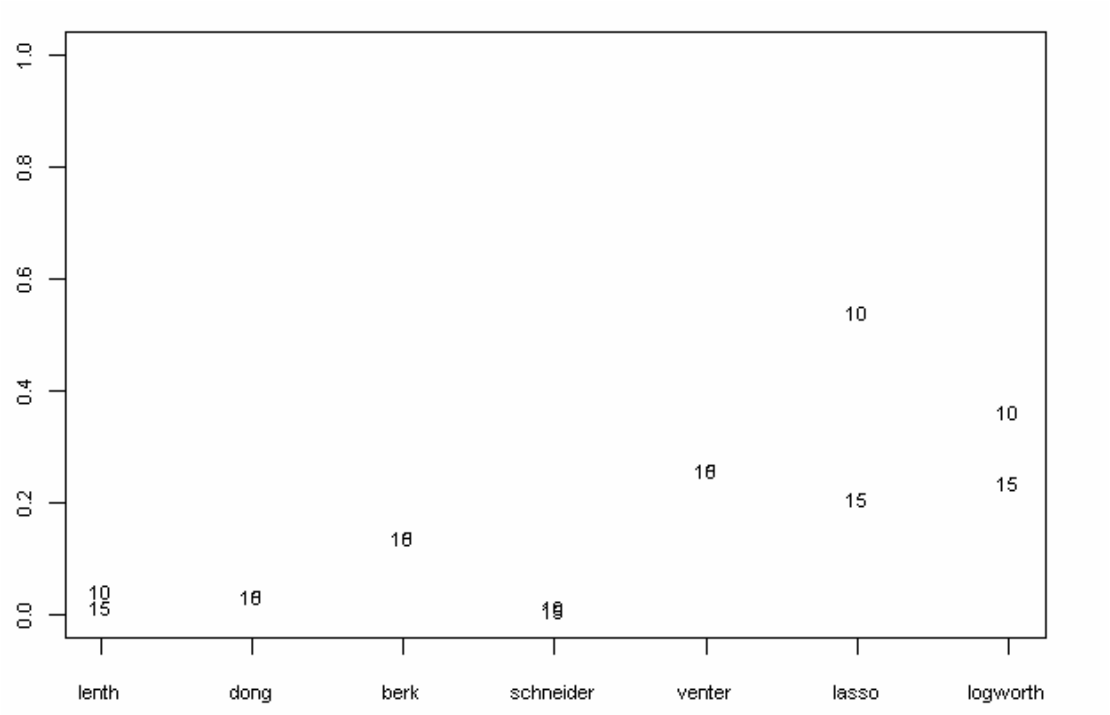

Figure 4.7.1: EER for situation with three equally sized large active effects. The numbers (10, 15) indicate the active effect mean-shift in number of standard deviations.

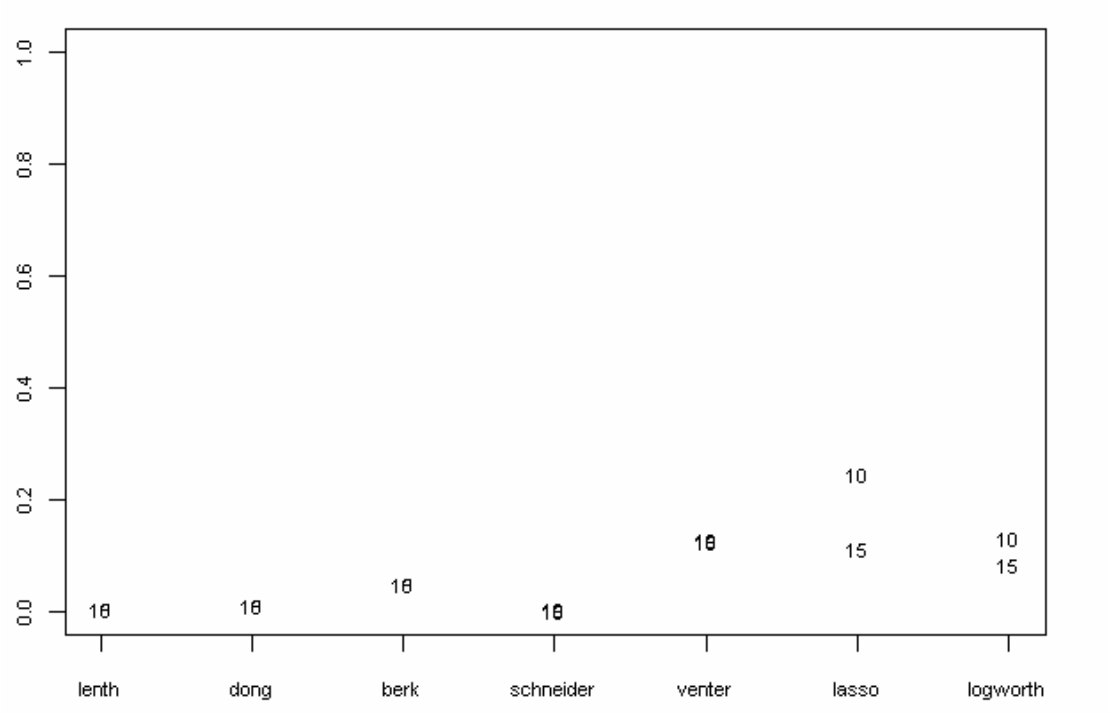

Figure 4.7.2: IER for situation with three equally sized large active effects. The numbers (10, 15) indicate the active effect mean-shift in number of standard deviations.

For three large effects, the Lasso had a slightly lower IER rate than when there was only one large effect present. The IER for Logworth and Venter were marginally worse, on the other hand. The other four methods once again proved to be the conservative ones, with IER close to or equal to zero.

None of the methods appeared to be altogether superior to the others in all of the different scenarios. The above results and implications for the application of these procedures will be discussed in the next chapter.

Chapter 5:
Discussion and Recommendations

5.1 Introduction

Some useful and interesting conclusions can be drawn in assessing the results obtained from the simulation study discussed in chapter 4. We will turn our attention now to these.

The research undertaken in this assignment also raised a number of questions, some of which still remain to be explored. This chapter is therefore concluded with a few suggestions of where further investigation on this topic might be aimed.

5.2 Discussion and recommendations

Of the seven methods included in the simulation study, none performed equally well under all circumstances. Neither did any technique arise as altogether superior to the others. Therefore it seems to be a matter of choosing the right tool for the task (as also noted by previous researchers). In the light of the results obtained from the simulation study, the following recommendations can be made:

- **no active effects** – in this situation (highly unlikely to occur in practice where 2^k factorial designs are typically used for screening experiments), the procedures proposed by Lenth and Schneider, Kasperski & Weissfeld perform the best. They are very conservative and have excellent EER and IER rates. Venter & Steel and Dong's methods also perform well, but the Logworth method is much too greedy and performs poorly.

- **small number of active effects** – Berk performs the best, with Dong and Schneider catching up as the effect magnitude is increased. Lenth's procedure seems to be too conservative for this kind of situation, notably failing to identify the smaller effects. The Logworth does slightly worse than Lenth in identifying the actual number of active effects, but still is much more liberal than any other method. The Lasso, also quite liberal, performs the worst by far.

- **large number of active effects** – the Logworth is the best procedure in terms of identifying the actual number of active effects (lowest EER), but still fails to be conservative. Lenth, Dong, Berk, Schneider, and the Lasso performs poorly, with Dong and Berk improving in performance only when (with six active effects) the effect size is increased to 6σ. Venter does well, and its performance increases as the number of active effects becomes larger.

- **small number of large effects** – Schneider emerged as the best procedure, with Lenth and Dong also performing excellently. Berk, Venter, and the Logworth does about equally well in identifying the actual number of active effects present, and the Lasso once again performs the worst of all. An interesting phenomenon can be observed regarding the Logworth method though: it seems that as the effect magnitude is increased, it tends to be less liberal, performing better overall (both its EER and IER rates decreasing).

These findings concur (where overlapping) with those made by Hamada & Balakrishnan in their thorough comparative simulation study (1998).

We can conclude that, for identifying active effects in unreplicated 2^k factorial experiments, the Lasso as variable selection technique did not turn out to be a significant addition to the methodological arsenal. The proposed Logworth procedure appears promising though, since it provides an improvement in performance over the

existing methods when the number of active effects becomes large relative to the total number of effects.

As has been noted before, in screening experiments, experimenters usually are not overly concerned with Type I errors (misclassifying inert effects as active). A bigger concern is Type II errors, which connotes misclassifying an active effect as inert, and dropping it from subsequent experimentation. Such errors will almost certainly ensure a failure to optimize the process in later stages (Hamada & Balakrishnan 1998: 34). The traditional, more conservative methods are designed to control the Type I error quite effectively, which consequently decreases their power and increases the probability of Type II errors. There is an acknowledged need for less conservative procedures (Hamada & Balakrishnan 1998: 37, 39) and the Logworth method seems to be a step in that direction.

While it may be argued that the Logworth procedure is still overly greedy, it showed a definite improvement as the number and magnitude of active effects were increased, proving to be a worthy competitor to the established, more conservative methods. It can also possibly be made more conservative by lowering the α value used in calculating the threshold logworth score, but this aspect was not fully explored in this assignment and will need further investigation.

Another advantage of the Logworth method is that its results can readily be displayed graphically. This is of much importance to practitioners (often non-statisticians), who will favour user-friendly and easily interpretable procedures (Hamada & Balakrishnan 1998: 33, 37). A simple graph of the logworth and threshold scores, as discussed in chapter 3, will meet this need.

5.3 Suggestions for further research

Though some promising new territory has been explored in this study, a number of points can be made to provide direction for further research. For one, the properties and merits of the proposed Logworth procedure should be investigated for a bigger range of scenarios, and also for some other popular sample sizes (e.g. 8 or 32 run designs), to altogether establish it as a worthy competitor to the more well-known techniques.

Secondly, more experimentation will be needed to find out how α, used to calculate the threshold logworth score, could possibly be varied to optimize the Logworth method under different circumstances.

Other modern variable selection methods can also be investigated, specifically with application to unreplicated factorial experiments. Some of these methods, like the Lasso, deals gracefully with situations where $p >> n$, and may be of use in identifying active effects in unreplicated 2^k factorial designs where p is nearly equal to n.

Appendices

A: 2^k Numerical example

B: *R* Software

C: Tables and figures

Example of a 2^3 factorial design

Finding $\hat{\beta}$ by the method of least squares will now be illustrated by borrowing an example given in Vining (1998: 374). It concerns a manufacturing operation for carbon-steel springs, in which a severe problem with cracks was experienced. Metallurgy theory at the time suggested that the cracking would be influenced by three factors:

- the temperature of the steel before quenching (Factor A)
- the amount of carbon in the formulation (Factor B)
- the temperature of the quenching oil (Factor C)

To investigate the effects of these process factors on the cracking of the springs, a 2^3 factorial experiment was set up. All springs were inspected for cracks, and the percentage of springs not showing signs of any cracking was chosen as the response variable. The design variables for the three factors were set up as follows:

$$x_1 = \begin{cases} -1 & \text{when steel temperature = 1450° F} \\ 1 & \text{when steel temperature = 1600° F} \end{cases}$$

$$x_2 = \begin{cases} -1 & \text{when percentage carbon = 0.50\%} \\ 1 & \text{when percentage carbon = 0.70\%} \end{cases}$$

$$x_3 = \begin{cases} -1 & \text{when oil temperature = 70° F} \\ 1 & \text{when oil temperature = 120° F} \end{cases}$$

The design is summarised (with the runs in *standard order*) in the table below:

Factors		A	B	C	AB	AC	BC	ABC
Design variables		x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
Runs	**Labels**							
1	(1)	-1	-1	-1	1	1	1	-1
2	*a*	1	-1	-1	-1	-1	1	1
3	*b*	-1	1	-1	-1	1	-1	1
4	*ab*	1	1	-1	1	-1	-1	-1
5	*c*	-1	-1	1	1	-1	-1	1
6	*ac*	1	-1	1	-1	1	-1	-1
7	*bc*	-1	1	1	-1	-1	1	-1
8	*abc*	1	1	1	1	1	1	1

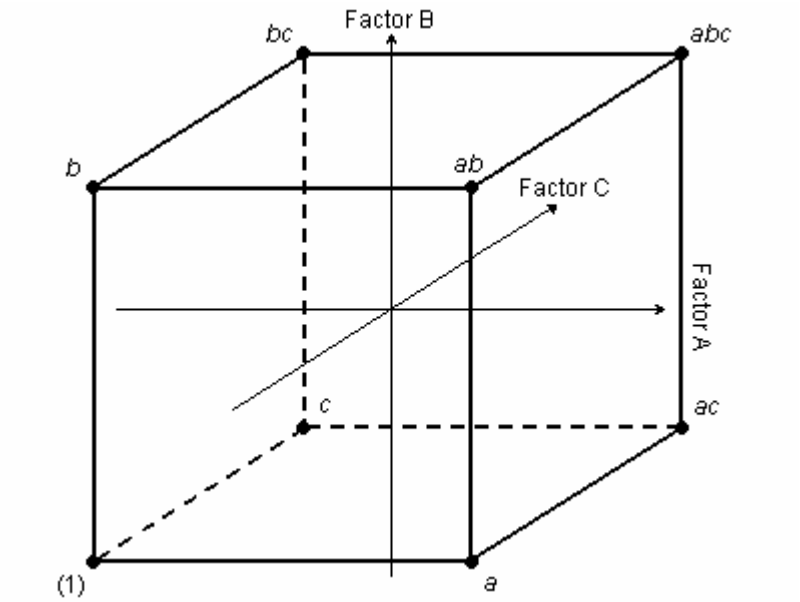

Figure A.1: Geometric representation of a 2^3 factorial experiment.

Geometrically, the design can be represented as a cube (*Figure A.1*), with each of the three dimensions representing one of the main factors (*A*, *B*, and *C*).

The design matrix and the vector of observed responses look as follows:

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \qquad Y = \begin{bmatrix} 67 \\ 79 \\ 61 \\ 75 \\ 59 \\ 90 \\ 52 \\ 87 \end{bmatrix}$$

For the design matrix, the columns of contrasts are preceded by a column of 1's, corresponding to the intercept term in a regression model. It can be seen that the contrasts are orthogonal: any column/row multiplied by any other column/row equals zero (Montgomery 2001: 231).

Then, because of using design variables equal to ± 1

$$X'X = \begin{bmatrix} 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \end{bmatrix}$$

All the diagonal elements of $X'X$ equals n, and all the off-diagonal elements equal zero, because of the orthogonality of the contrasts.

Then

$$X'Y = \begin{bmatrix} 570 \\ 92 \\ -20 \\ 6 \\ 6 \\ 40 \\ 0 \\ 2 \end{bmatrix}$$

and the normal equations can be found as

$$X'X\beta = \begin{bmatrix} 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_{12} \\ \beta_{13} \\ \beta_{23} \\ \beta_{123} \end{bmatrix} = \begin{bmatrix} 570 \\ 92 \\ -20 \\ 6 \\ 6 \\ 40 \\ 0 \\ 2 \end{bmatrix} = X'Y$$

which becomes

$$\begin{aligned}
8\beta_0 &= 570\\
8\beta_1 &= 92\\
8\beta_2 &= -20\\
8\beta_3 &= 6\\
8\beta_{12} &= 6\\
8\beta_{13} &= 40\\
8\beta_{23} &= 0\\
8\beta_{123} &= 2
\end{aligned}$$

The inverse of $X'X$ is given by

$$(X'X)^{-1} = \begin{bmatrix}
0.125 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0.125 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0.125 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0.125 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0.125 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0.125 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0.125 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.125
\end{bmatrix}$$

and then the estimated regression coefficients can be calculated as

$$\hat{\beta} = (X'X)^{-1} X'Y = \begin{bmatrix}
71.25\\
11.5\\
-2.5\\
0.75\\
0.75\\
5\\
0\\
0.25
\end{bmatrix}$$

It can be seen that the estimated regression coefficients are equal to exactly half the estimated effects (Vining 1998: 375). This is no surprise, for $\hat{\beta}_i$ is the estimated change in the response for a change equal to one in the *i*th factor, while the *i*th factor effect is measured as the change in the response when the *i*th factor is varied from -1 to +1 (i.e. a change of two units of measurement).

Computation in R

All data analysis and simulation work was done in *R* version 2.2.0 (R Development Core Team: 2005). The add-on packages ***lars*** (Hastie & Efron: 2004) and ***lasso2*** (Lokhorst et al: 2006) were downloaded from the *R* website (www.R-project.org) and employed in the routines for computing the Lasso and Logworth solutions.

R is an open-source language with numerous contributors. It may be freely downloaded and shared. It is an interpreted programming language, which slows down its processing speed, but this disadvantage is outweighed by several data handling features specifically designed for statistical computing.

Simulating response data

```
simsamp<-function(design.mat,mu.vec,sig.mat,nsamp=100)
{
# design.mat: design matrix of the experiment
# mu.vec: vector of mean effects
# sig.mat: covariance matrix of the effects
# nsamp: number of samples taken

      library(MASS)
      n<-nrow(design.mat)
      p<-length(mu.vec)
      effects<-matrix(NA, ncol=p,nrow=nsamp)
      response<-matrix(NA, ncol=n,nrow=nsamp)
      contrast<-matrix(NA, ncol=p,nrow=nsamp)
      for(i in 1:nsamp){
            effects[i,]<-mvrnorm(n=1,mu=mu.vec,Sigma=sig.mat)
            response.vec<-rep(NA,times=n)
            for(j in 1:n){
                  response.vec[j]<-design.mat[j,]%*%effects[i,]
            }
            response[i,]<-response.vec
            contrast.vec<-rep(NA,times=p)
            for(k in 1:p){
                  contrast.vec[k]<-design.mat[,k]%*%response[i,]
            }
            contrast[i,]<-contrast.vec
      }
      return(list(response=response,contrast=contrast))
}
```

Lenth's method

```
lenth<-function(designmat,response)
{
# Lenth's method for testing the significance of effects in
# unreplicated factorial experiments, as in Montgomery p254

      m<-ncol(designmat)
      contr<-vector("numeric",length=m)
      for(i in 1:m){
            contr[i]<-abs(designmat[,i]%*%matrix(response,ncol=1))
      }

      s0<-1.5*median(contr)
      normcontr<-NULL
      for(i in 1:m){
            if(contr[i]<(2.5*s0)){normcontr<-c(normcontr,contr[i])}
      }

      PSE<-1.5*median(normcontr)
      d<-m/3
      library(stats)
      ME<-qt(0.025,df=d,lower.tail=F)*PSE
      delta<-1-(1+0.95^(1/m))/2
      SME<-qt(delta,d,lower.tail=F)*PSE
      ind<-NULL
      for(i in 1:length(contr)){
            if(contr[i]>SME){ind<-c(ind,"*")}
            else {ind<-c(ind," ")}
      }
      finalform<-data.frame(contrast=contr,significant=ind)
      rownames(finalform)<-colnames(designmat)

      return(list(PSE=PSE,ME=ME,SME=SME,report=finalform))
}
```

Dong's method

```
dong<-function(designmat, response)
{
# Dong's method for testing the significance of effects in
# unreplicated factorial experiments, as in Hamada & Balakrishnan p12

      k<-ncol(designmat)
      abscontr<-vector("numeric",length=k)
      for(i in 1:k){
            abscontr[i]<-abs(designmat[,i]%*%matrix(response,ncol=1))
      }

      s0<-1.5*median(abscontr)
      inactive<-NULL
      for(i in 1:k){
            if(abscontr[i]<(2.5*s0)){
                  inactive<-c(inactive,abscontr[i])}
      }

      m<-length(inactive)
      s.dong<-sqrt(sum(inactive^2)/m)
```

```
      gamma<-(1+0.95^(1/k))/2
      ME<-qt(0.975,df=m)*s.dong
      SME<-qt(p=gamma,df=m)*s.dong

      ind<-NULL
      for(i in 1:k){
            if(abscontr[i]>SME){ind<-c(ind,"*")}
            else {ind<-c(ind," ")}
      }
      finalform<-data.frame(contrast=abscontr,significant=ind)
      rownames(finalform)<-colnames(designmat)

      return(list(s.dong=s.dong,report=finalform))
}
```

Berk & Picard's method

```
berk<-function(designmat,response)
{
# Berk & Picard's method for testing the significance of effects in
# unreplicated factorial experiments, as in Berk & Picard (1991) p80

      k<-ncol(designmat)
      ss<-as.vector(((t(designmat)%*%response)^2)/(k+1))
      ss.sort<-sort(ss)
      m<-round(k*0.6)
      mse<-sum(ss.sort[1:m])/m
      Bvals<-ss/mse

      if(k==7){critval<-23.76}
      else if(k==15){critval<-18.93}
      else if(k==31){critval<-17.46}
      else {cat("Critical value unknown: Reference Berk & Picard
            (1991) p81\n")}

      ind<-NULL
      for(i in 1:k){
            if(Bvals[i]>critval){ind<-c(ind,"*")}
            else {ind<-c(ind," ")}
      }
      finalform<-data.frame(SS=ss,significant=ind)
      rownames(finalform)<-colnames(designmat)

      return(list(mse=mse,report=finalform))
}
```

Schneider, Kasperski & Weissfeld's method

```
Schneider<-function(designmat,response)
{
# Schneider, Kasperski & Weissfeld's method for testing the
# significance of effects in unreplicated factorial experiments, as
# in Schneider, Kasperski & Weissfeld (1993) p19-20,25-26

      k<-ncol(designmat)
      n<-round(k/2)
      abscontr<-as.vector(abs(t(designmat)%*%response))
      abscontr.sort<-sort(abscontr)
      M<-abscontr.sort[n]     #n-th smallest unsigned contrast
      sn<-sd(abscontr.sort[1:n])      #std.dev of the n smallest
                                          unsigned contrasts
      sigma<-sn
      r<-k-n
      repeat {
            H<-dnorm(M/sigma)/(1-pnorm(M/sigma))    #hazard function
            A<-r/n*H*M
            newsigma<-A/2+sqrt(A^2/4+sn^2)
            if(abs(sigma-newsigma)<0.000001){sigma<-newsigma; break}
            else {sigma<-newsigma}
      }
      tvals<-abscontr/sigma

      u<-qnorm(n/k)
      gamma<-1/(2*n/k-u*dnorm(u)+u^2*(k*dnorm(u)/r-u))
      alpha<-1-0.95^(1/k)     #for computing the outer limits; for
                                    inner limits, use alpha = 0.05
      z2<-(qnorm(1-alpha/2))^2
      tau<-sqrt(z2/(1-gamma/k*z2))

      ind<-NULL
      for(i in 1:k){
            if(tvals[i]>tau){ind<-c(ind,"*")}
            else {ind<-c(ind," ")}
      }
      finalform<-data.frame(contrasts=abscontr,significant=ind)
      rownames(finalform)<-colnames(designmat)

      return(list(sigma=sigma,tau=tau,report=finalform))
}
```

Venter & Steel's method

```
venter
function (designmat,response,l=3,B=1000,critval=NA)
{
# Venter & Steel's method for testing the significance of effects in
# unreplicated factorial experiments, as in Venter & Steel (1996)
# p162-163

      k<-ncol(designmat)
      critvec7<-c(0.0087,0.0104,0.0130,0.0174,0.0260)
      critvec15<-c(0.0038,0.0041,0.0045,0.0050,0.0055,0.0061,0.0069,
                        0.0079,0.0092,0.0110,0.0137,0.0179)
      if(is.na(critval)){
            if(k==7){critval<-critvec7[l]}
            else if(k==15){critval<-critvec15[l]}
            else {cat("Percentile of S distribution not specified\n")
                  return()}
      }
      abscontr<-as.vector(abs(t(designmat)%*%response))
      abscontr.sort<-sort(abscontr)

# Calculating V-ratios
      V<-rep(NA,times=(k-1))
      for(i in 1:(k-1)){
            g<-sqrt(sum(abscontr.sort[1:i]^2)/i)
            V[i]<-abscontr.sort[i+1]/g
      }

# Calculating p-values for V-ratios
      Tmat<-matrix(abs(rnorm(n=B*k)),ncol=k)
      TVmat<-matrix(NA,nrow=B,ncol=(k-1))
      for(b in 1:B){
            for(i in 1:(k-1)){
                  Tmat[b,]<-sort(Tmat[b,])
                  g<-sqrt(sum(Tmat[b,1:i]^2)/i)
                  TVmat[b,i]<-Tmat[b,i+1]/g
            }
      }
      Vcount<-rep(0,times=(k-1))
      for(i in 1:(k-1)){
            for(b in 1:B){
                  if(TVmat[b,i]>V[i]){Vcount[i]<-Vcount[i]+1}
            }
      }
      pvals<-Vcount/B

      q<-k
      for(i in 1:(k-1)){
            if(pvals[i]<=critval){q<-i; break}
      }

      return(list(p.values=pvals,activenum=(k-q)))
}
```

Lasso method

```
tibshirani<-function(designmat,response)
{
# The Lasso method for testing the significance of effects in
# unreplicated factorial experiments, as in Tibshirani (1996)

      library(lars)
      library(lasso2)
      totaldata<-data.frame(designmat,response)
      t.constr<-min(cv.lars(x=as.matrix(designmat),y=response,
                        K=nrow(designmat),plot.it=FALSE,se=FALSE)$cv)
      lasso.coef<-coef(l1ce(response~0+A*B*C*D,data=totaldata,
                                    absolute.t=TRUE,bound=t.constr))

      k<-length(lasso.coef)
      ind<-NULL
      for(i in 1:k){
            if(lasso.coef[i]!=0){ind<-c(ind,"*")}
            else {ind<-c(ind," ")}
      }
      finalform<-data.frame(coef=lasso.coef,significant=ind)
      rownames(finalform)<-colnames(designmat)

      return(list(t.constraint=t.constr,report=finalform))
}
```

Note: This routine makes use of the ***lars*** and ***lasso2*** packages, which can be downloaded from www.R-project.org.

Logworth method

Simulation version (shorter runtime)

```
lars.sequence<-function(coefmat)
{
      p<-ncol(coefmat)
      n<-nrow(coefmat)
      sequence<-NULL
      included<-vector("logical",length=p)
      for(i in 2:n){
            for(j in 1:p){
                  if(!included[j]&&(coefmat[i,j]!=0)){
                        sequence<-append(sequence,j)
                        included[j]<-TRUE
                  }
            }
      }
      return(sequence)
}
```

```
findopt<-function(datavec,critval)
{
      n<-length(datavec)
      optpoints<-NULL
      for(i in 1:n){
            if(i==1){

                  if((datavec[i]>critval)&&(datavec[i]>datavec[i+1]))
                        {optpoints<-append(optpoints,i)}
            }
            else if(i==n){
                  if((datavec[i]>critval)&&(datavec[i]>datavec[i-1]))
                        {optpoints<-append(optpoints,i)}
            }
            else if((datavec[i]>critval)&&((datavec[i]>datavec[i-
                        1])&&(datavec[i]>datavec[i+1])))
                              {optpoints<-append(optpoints,i)}

      }
      if(is.null(optpoints)){optpoints<-0}
      return(optpoints)
}


logworth<-function(designmat,response,alpha=0.05)
{
# Uses the lasso sequence of models and selects one by comparing
# their logworth (-ln(pvalue))

      library(lars)
      response<-response-mean(response)
      p<-ncol(designmat)
      n<-length(response)
      model.pvals<-rep(NA,times=(p-1))
      sequence<-lars.sequence(coef.lars(lars(x=designmat,
                                             y=response)))
      for(i in 1:(p-1)){
            y<-matrix(response,ncol=1)
            x<-designmat[,sequence[1:i],drop=F]
            df1<-ncol(x)
            df2<-n-df1-1
            B<-solve(t(x)%*%x)%*%t(x)%*%y
            term<-t(B)%*%t(x)%*%y
            SSE<-t(y)%*%y-term
            SSR<-term-n*(mean(y)^2)
            fstat<-(SSR/df1)/(SSE/df2)
            model.pvals[i]<-1-pf(fstat,df1,df2)
      }

      log.pvals<- -log(model.pvals)
      critval<- -log(alpha)
      activenum<-findopt(log.pvals,critval)[1]
      return(list(activenum=activenum,min.log.pvals=log.pvals))
}
```

Note: This routine makes use of the ***lars*** package, which can be downloaded from www.R-project.org.

Exact version

```
combin<-function(r,n,vek=1:n)
{
      if(r<=0){return(NULL)}
      else if(r>=n){return(vek[1:n])}
      else {return(rbind(cbind(vek[1],Recall(r-1,n-1,
                              vek[-1])),Recall(r,n-1,vek[-1])))}
}
```

Note: The above routine was developed by Van der Merwe (2004).

```
logworth.exact<-function(designmat,response,alpha=0.05,graph=TRUE)
{
# Fits models for all combinations of variable subsets, and compares
# the best model of each subset size by comparing -ln(pvalue)

      p<-ncol(designmat)
      response<-response-mean(response)
      results<-lapply(rep(NA,times=(p-1)),"c")
      model.pvals<-rep(NA,times=(p-1))
      varnames<-colnames(designmat)
      for(i in 1:(p-1)){
            varcomb<-combin(r=i,n=p)
            if(i==p){varcomb<-matrix(varcomb,nrow=1)}
            ncomb<-nrow(varcomb)
            minpval<-10
            for(k in 1:ncomb){
                  kdata<-data.frame(matrix(response,ncol=1),
                                          designmat[,varcomb[k,]])
                  colnames(kdata)<-c("response",
                                          varnames[varcomb[k,]])
                  kmodel<-lm(formula=as.formula(paste("response~",
                                    paste(varnames[varcomb[k,]],
                                    collapse="+"))),data=kdata)
                  fstat<-summary(kmodel)$fstatistic
                  pval<-1-pf(fstat[1],fstat[2],fstat[3])
                  if(pval<minpval){minpval<-pval
                                    results[[i]]<-coef(kmodel)}
            }
            model.pvals[i]<-minpval
      }
      log.pvals<- -log(model.pvals)
      critval<- -log(alpha)
            if(graph){
            par(pch=20)
            plot(1:(p-1),log.pvals,xlab="number of variables",
                  ylab="-ln(p-value)",main="Logworth",
                  ylim=c(0,max(log.pvals)))
            lines(1:(p-1),log.pvals)
            abline(h=critval,lty=3)
      }
      activenum<-matrix(findopt(log.pvals,critval),nrow=1)[1,1]
      return(list(activenum=activenum,model=results[[activenum]],
                                          min.log.pvals=log.pvals))
}
```

Note: This routine makes use of the ***lars*** package, which can be downloaded from www.R-project.org.

Simulation routines

```
count.active<-function(datavec,indnum)
{
      n<-length(datavec)
      results<-rep(0,times=indnum)
      for(i in 1:n){
            results[datavec[i]+1]<-results[datavec[i]+1]+1
      }
      return(results)
}


simulasie<-function (designmat=x16,mu=mu16,sigma=sigma16,nsim=100,
                     shownum=8,active,
            outputpath="C:/Stats/Eksperimenteel/Simulasie/Output")
{
      begin.time<-Sys.time()
      simresponse<-simsamp(design.mat=designmat,mu.vec=mu,
                           sig.mat=sigma,nsamp=nsim)$response
      write.table(simresponse,
            file=paste(outputpath,"/simresponse.txt",sep=""),
            quote=FALSE,row.names=FALSE,col.names=FALSE,sep="\t")

# Lenth's method

      signif<-matrix(NA,nrow=nsim,ncol=ncol(designmat))
      for(i in 1:nsim){
            signif[i,]<-lenth(designmat,
                           simresponse[i,])$report[,"significant"]
        }
      colnames(signif)<-colnames(designmat)
      lenth.active<-apply(signif-1,1,sum)
      write(lenth.active,
                  file=paste(outputpath,"/lenth.out.txt",sep=""),
                  ncolumns=10,sep="\t")

# Dong's method

      signif<-matrix(NA,nrow=nsim,ncol=ncol(designmat))
      for(i in 1:nsim){
            signif[i,]<-dong(designmat,
                           simresponse[i,])$report[,"significant"]
      }
      colnames(signif)<-colnames(designmat)
      dong.active<-apply(signif-1,1,sum)
      write(dong.active,
                  file=paste(outputpath,"/dong.out.txt",sep=""),
                  ncolumns=10,sep="\t")
```

```
# Berk & Picard's method

      signif<-matrix(NA,nrow=nsim,ncol=ncol(designmat))
      for(i in 1:nsim){
            signif[i,]<-berk(designmat,
                              simresponse[i,])$report[,"significant"]
      }
      colnames(signif)<-colnames(designmat)
      berk.active<-apply(signif-1,1,sum)
      write(berk.active,
                  file=paste(outputpath,"/berk.out.txt",sep=""),
                  ncolumns=10,sep="\t")

# Schneider, Kasperski & Weissfeld's method

      signif<-matrix(NA,nrow=nsim,ncol=ncol(designmat))
      for(i in 1:nsim){
            signif[i,]<-schneider(designmat,
                              simresponse[i,])$report[,"significant"]
      }
      colnames(signif)<-colnames(designmat)
      schneider.active<-apply(signif-1,1,sum)
      write(schneider.active,
                  file=paste(outputpath,"/schneider.out.txt",sep=""),
                  ncolumns=10,sep="\t")

# Venter & Steel's method

      venter.active<-rep(NA,times=nsim)
      for(i in 1:nsim){
            venter.active[i]<-venter(designmat,
                              simresponse[i,],B=50,l=5)$activenum
      }
      write(venter.active,
                  file=paste(outputpath,"/venter.out.txt",sep=""),
                  ncolumns=10,sep="\t")

# Lasso method

      signif<-matrix(NA,nrow=nsim,ncol=ncol(designmat))
      for(i in 1:nsim){
            signif[i,]<-tibshirani(designmat,
                              simresponse[i,])$report[,"significant"]
      }
      colnames(signif)<-colnames(designmat)
      lasso.active<-apply(signif-1,1,sum)
      write(lasso.active,
                  file=paste(outputpath,"/lasso.out.txt",sep=""),
                  ncolumns=10,sep="\t")

# Logworth method

      logworth.active<-rep(NA,times=nsim)
      for(i in 1:nsim){
            logworth.active[i]<-logworth(designmat,
                              simresponse[i,],alpha=0.05)$activenum
      }
      write(logworth.active,
                  file=paste(outputpath,"/logworth.out.txt",sep=""),
                  ncolumns=10,sep="\t")
```

```
      end.time<-Sys.time()
      cat(nsim,"simulations took:",end.time-begin.time,"\n")

      lenth.tab<-append(count.active(lenth.active,
                  indnum=(max(lenth.active)+1)),rep(0,times=50))/nsim
      dong.tab<-append(count.active(dong.active,
                  indnum=(max(dong.active)+1)),rep(0,times=50))/nsim
      berk.tab<-append(count.active(berk.active,
                  indnum=(max(berk.active)+1)),rep(0,times=50))/nsim
      schneider.tab<-append(count.active(schneider.active,
            indnum=(max(schneider.active)+1)),rep(0,times=50))/nsim
      venter.tab<-append(count.active(venter.active,
            indnum=(max(venter.active)+1)),rep(0,times=50))/nsim
      lasso.tab<-append(count.active(lasso.active,
                  indnum=(max(lasso.active)+1)),rep(0,times=50))/nsim
      logworth.tab<-append(count.active(logworth.active,
            indnum=(max(logworth.active)+1)),rep(0,times=50))/nsim
      results<-data.frame(effects=0:(shownum-1),
            lenth=lenth.tab[1:shownum],dong=dong.tab[1:shownum],
            berk=berk.tab[1:shownum],
            schneider=schneider.tab[1:shownum],
            venter=venter.tab[1:shownum],lasso=lasso.tab[1:shownum],
            logworth=logworth.tab[1:shownum])
      write.table(results,
                        file=paste(outputpath,"/results.txt",sep=""),
                        quote=F,row.names=F,sep="\t")

      EER<-1-results[active+1,,drop=F]
      IER<-rep(0,times=7)
      for(i in (active+2):shownum){
            IER<-IER+(results[i,"effects"]/(nrow(designmat)-
                                    (active+1)))*results[i,,drop=F]
      }
      error.rates<-rbind(EER,IER)[,-1]
      rownames(error.rates)<-c("EER","IER")
      write.table(error.rates,
                  file=paste(outputpath,"/error.rates.txt",sep=""),
                  quote=F,sep="\t")

      return(list(results=results,error.rates=error.rates))
}
```

```
globalsim<-function (designmat,mumat,sigma,N,
                  dirpath="C:/Stats/Eksperimenteel/Simulasie/Output")
{
# designmat: design matrix
# mumat: matrix of which the rows contain different scenarios for the
#        average factor effects
# sigma: covariance matrix of the factor effects
# N: number of simulations per scenario
# outputpath: directory path to which output will be written

      n.scen<-nrow(mumat)
      for(i in 1:n.scen){
            outputdir<-paste(dirpath,"/Scenario",i,sep="")
            write(mumat[i,],
                        file=paste(outputdir,"/mu.vec.txt",sep=""),
                        ncolumns=10)
            numactive<-length(mumat[i,mumat[i,]!=0])
            simulasie(designmat,mumat[i,],sigma,N,
                        shownum=ncol(mumat),active=numactive,
                        outputpath=outputdir)
      }
}
```

No active effects

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.979	0.920	0.552	0.953	0.845	0.712	0.002
	1	0.013	0.058	0.259	0.038	0.012	0.000	0.035
	2	0.004	0.014	0.116	0.007	0.014	0.001	0.038
	3	0.002	0.006	0.053	0.002	0.015	0.004	0.058
	4	0.001	0.002	0.016	0.001	0.014	0.006	0.082
	5	0.001	0.000	0.005	0.000	0.014	0.010	0.113
	6	0.000	0.000	0.000	0.000	0.017	0.013	0.138
	7	0.000	0.000	0.000	0.000	0.016	0.019	0.148
	8	0.000	0.000	0.000	0.000	0.017	0.019	0.145
	9	0.000	0.000	0.000	0.000	0.018	0.022	0.117
	10	0.000	0.000	0.000	0.000	0.018	0.021	0.074
	11	0.000	0.000	0.000	0.000	0.000	0.026	0.036
	12	0.000	0.000	0.000	0.000	0.000	0.033	0.011
	13	0.000	0.000	0.000	0.000	0.000	0.049	0.001
	14	0.000	0.000	0.000	0.000	0.000	0.067	0.001
EER		0.021	0.080	0.448	0.047	0.155	0.288	0.998
IER		0.002	0.008	0.049	0.004	0.060	0.207	0.437

Table 4.1: Experimental and individual error rates of the seven different methods when there are no active effects (off-the-shelf performance). Cells contain the proportions effects declared active by the specific method for 10 000 simulated data sets, using an unreplicated 2^4 factorial design (16 runs).

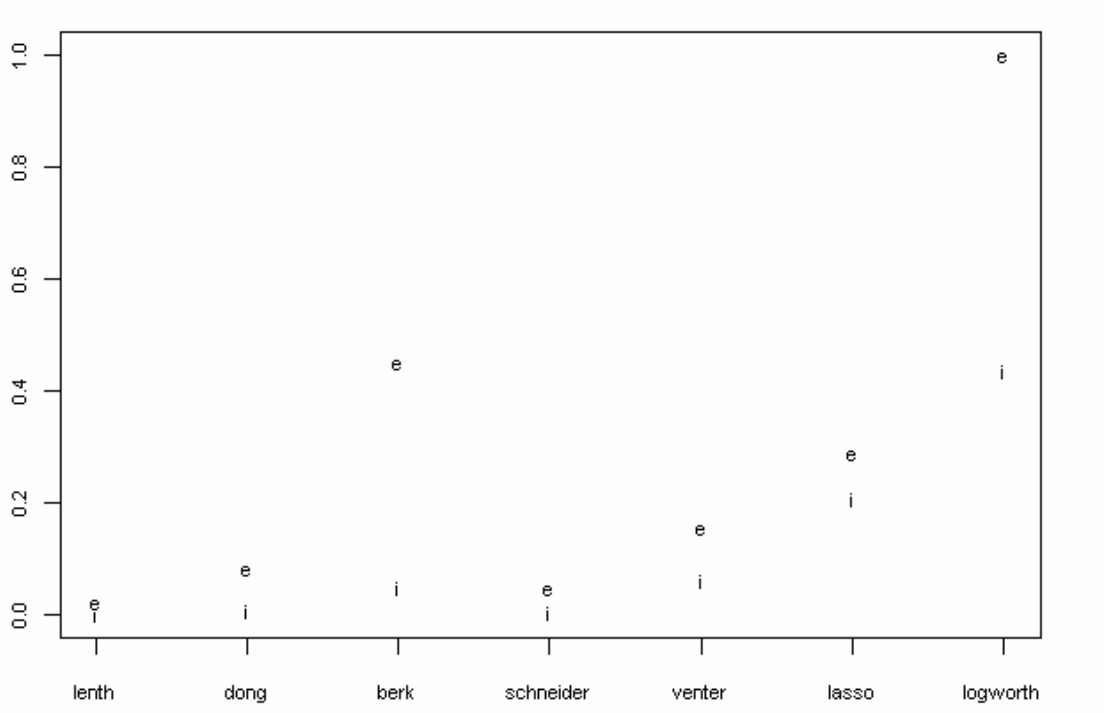

Figure 4.1: Experimental and individual error rates (denoted by "e" and "i", respectively) of the seven different methods when there are no active effects (off-the-shelf performance).

Small number of active effects

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.980	0.914	0.542	0.955	0.845	0.738	0.002
	1	0.013	0.064	0.260	0.036	0.015	0.000	0.034
	2	0.005	0.015	0.125	0.008	0.015	0.002	0.039
	3	0.001	0.004	0.050	0.001	0.014	0.003	0.057
	4	0.001	0.002	0.018	0.000	0.016	0.006	0.083
	5	0.000	0.001	0.005	0.000	0.014	0.009	0.121
	6	0.000	0.000	0.000	0.000	0.013	0.012	0.141
	7	0.000	0.000	0.000	0.000	0.015	0.016	0.150
	8	0.000	0.000	0.000	0.000	0.015	0.018	0.143
	9	0.000	0.000	0.000	0.000	0.019	0.024	0.111
	10	0.000	0.000	0.000	0.000	0.018	0.022	0.072
	11	0.000	0.000	0.000	0.000	0.000	0.025	0.035
	12	0.000	0.000	0.000	0.000	0.000	0.029	0.011
	13	0.000	0.000	0.000	0.000	0.000	0.038	0.002
	14	0.000	0.000	0.000	0.000	0.000	0.059	0.001
	EER	0.987	0.936	0.740	0.964	0.985	1.000	0.966
	IER	0.001	0.004	0.036	0.001	0.062	0.199	0.463

Table 4.2.1: Experimental and individual error rates for a single active effect with a mean-shift of 1σ .

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.953	0.825	0.419	0.900	0.785	0.740	0.001
	1	0.035	0.140	0.338	0.084	0.047	0.001	0.055
	2	0.007	0.025	0.154	0.013	0.024	0.003	0.055
	3	0.002	0.007	0.060	0.002	0.020	0.005	0.070
	4	0.002	0.002	0.023	0.001	0.021	0.008	0.092
	5	0.000	0.001	0.005	0.000	0.016	0.010	0.124
	6	0.000	0.001	0.001	0.000	0.015	0.014	0.136
	7	0.000	0.000	0.000	0.000	0.015	0.018	0.146
	8	0.000	0.000	0.000	0.000	0.016	0.019	0.126
	9	0.000	0.000	0.000	0.000	0.021	0.021	0.098
	10	0.000	0.000	0.000	0.000	0.020	0.022	0.058
	11	0.000	0.000	0.000	0.000	0.000	0.023	0.029
	12	0.000	0.000	0.000	0.000	0.000	0.029	0.007
	13	0.000	0.000	0.000	0.000	0.000	0.038	0.002
	14	0.000	0.000	0.000	0.000	0.000	0.049	0.000
	EER	0.965	0.860	0.662	0.916	0.953	0.999	0.945
	IER	0.002	0.006	0.044	0.003	0.070	0.190	0.431

Table 4.2.2: Experimental and individual error rates for a single active effect with a mean-shift of 2σ .

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.753	0.291	0.065	0.511	0.429	0.641	0.000
	1	0.226	0.648	0.599	0.462	0.363	0.007	0.217
	2	0.015	0.046	0.221	0.024	0.046	0.009	0.133
	3	0.004	0.010	0.083	0.003	0.029	0.011	0.131
	4	0.002	0.004	0.026	0.001	0.021	0.013	0.120
	5	0.000	0.001	0.005	0.000	0.019	0.017	0.115
	6	0.000	0.000	0.001	0.000	0.018	0.022	0.101
	7	0.000	0.000	0.000	0.000	0.017	0.023	0.076
	8	0.000	0.000	0.000	0.000	0.019	0.029	0.053
	9	0.000	0.000	0.000	0.000	0.019	0.026	0.032
	10	0.000	0.000	0.000	0.000	0.021	0.030	0.016
	11	0.000	0.000	0.000	0.000	0.000	0.030	0.005
	12	0.000	0.000	0.000	0.000	0.000	0.038	0.002
	13	0.000	0.000	0.000	0.000	0.000	0.047	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.058	0.000
	EER	0.774	0.352	0.401	0.538	0.637	0.994	0.783
	IER	0.004	0.010	0.059	0.004	0.080	0.246	0.271

Table 4.2.3: Experimental and individual error rates for a single active effect with a mean-shift of 4σ .

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.354	0.017	0.001	0.118	0.089	0.518	0.000
	1	0.627	0.918	0.656	0.855	0.699	0.031	0.408
	2	0.013	0.051	0.227	0.022	0.050	0.016	0.206
	3	0.003	0.010	0.083	0.004	0.029	0.019	0.135
	4	0.002	0.004	0.025	0.001	0.021	0.024	0.101
	5	0.001	0.001	0.007	0.000	0.020	0.025	0.067
	6	0.000	0.000	0.001	0.000	0.019	0.024	0.042
	7	0.000	0.000	0.000	0.000	0.018	0.027	0.026
	8	0.000	0.000	0.000	0.000	0.016	0.033	0.010
	9	0.000	0.000	0.000	0.000	0.021	0.034	0.004
	10	0.000	0.000	0.000	0.000	0.020	0.034	0.001
	11	0.000	0.000	0.000	0.000	0.000	0.042	0.000
	12	0.000	0.000	0.000	0.000	0.000	0.048	0.000
	13	0.000	0.000	0.000	0.000	0.000	0.056	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.072	0.000
	EER	0.373	0.082	0.344	0.145	0.301	0.969	0.592
	IER	0.004	0.011	0.060	0.004	0.080	0.308	0.151

Table 4.2.4: Experimental and individual error rates for a single active effect with a mean-shift of 6σ .

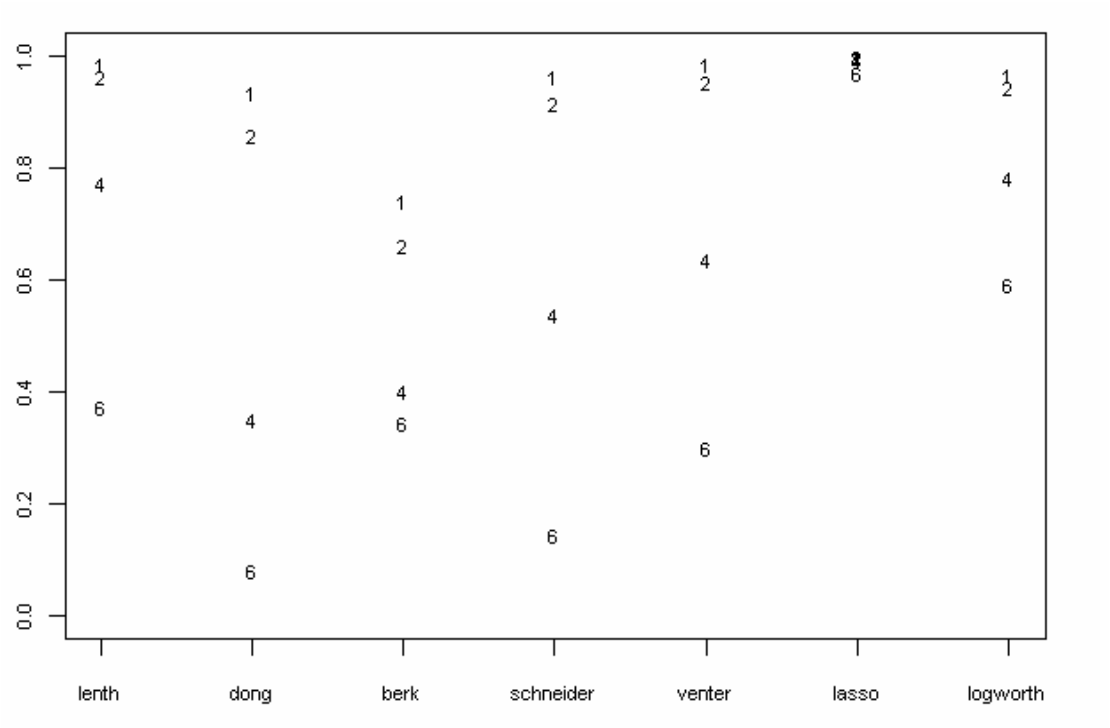

Figure 4.2.1: EER for situation with a single active effect present. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

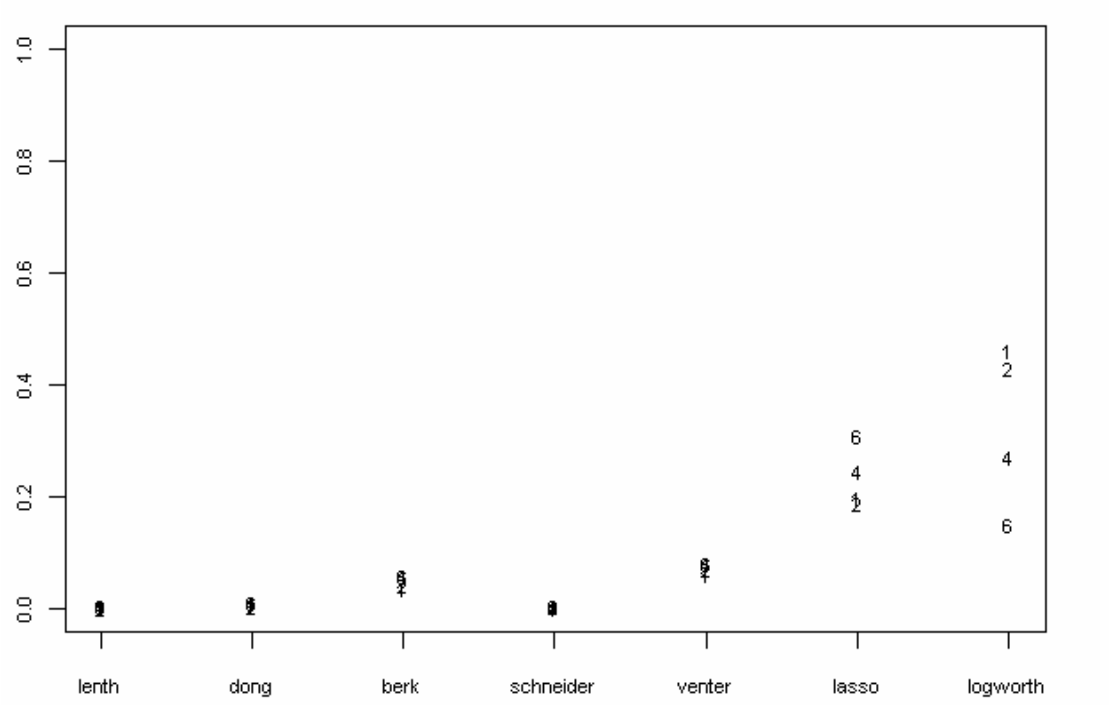

Figure 4.2.2: IER for situation with a single active effect present. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.974	0.898	0.523	0.948	0.839	0.772	0.002
	1	0.016	0.080	0.273	0.042	0.017	0.000	0.037
	2	0.006	0.016	0.128	0.009	0.014	0.001	0.041
	3	0.002	0.005	0.054	0.002	0.016	0.003	0.060
	4	0.001	0.001	0.017	0.000	0.013	0.005	0.084
	5	0.000	0.001	0.004	0.000	0.016	0.008	0.115
	6	0.000	0.000	0.001	0.000	0.015	0.013	0.136
	7	0.000	0.000	0.000	0.000	0.019	0.016	0.150
	8	0.000	0.000	0.000	0.000	0.016	0.017	0.142
	9	0.000	0.000	0.000	0.000	0.016	0.019	0.114
	10	0.000	0.000	0.000	0.000	0.019	0.019	0.069
	11	0.000	0.000	0.000	0.000	0.000	0.023	0.036
	12	0.000	0.000	0.000	0.000	0.000	0.028	0.012
	13	0.000	0.000	0.000	0.000	0.000	0.030	0.002
	14	0.000	0.000	0.000	0.000	0.000	0.046	0.000
	EER	0.998	0.995	0.946	0.999	0.984	0.997	0.940
	IER	0.001	0.001	0.008	0.000	0.068	0.197	0.516

Table 4.3.1: Experimental and individual error rates for a situation with three equal-sized active effects (mean-shift of 1σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.950	0.829	0.375	0.900	0.760	0.808	0.001
	1	0.032	0.114	0.300	0.075	0.025	0.001	0.036
	2	0.012	0.038	0.194	0.019	0.038	0.002	0.049
	3	0.004	0.014	0.092	0.005	0.033	0.004	0.071
	4	0.002	0.003	0.031	0.001	0.026	0.006	0.104
	5	0.001	0.001	0.008	0.000	0.020	0.010	0.130
	6	0.000	0.000	0.001	0.000	0.020	0.014	0.148
	7	0.000	0.000	0.000	0.000	0.018	0.015	0.146
	8	0.000	0.000	0.000	0.000	0.020	0.016	0.141
	9	0.000	0.000	0.000	0.000	0.020	0.018	0.092
	10	0.000	0.000	0.000	0.000	0.019	0.017	0.053
	11	0.000	0.000	0.000	0.000	0.000	0.019	0.023
	12	0.000	0.000	0.000	0.000	0.000	0.022	0.006
	13	0.000	0.000	0.000	0.000	0.000	0.024	0.001
	14	0.000	0.000	0.000	0.000	0.000	0.027	0.000
	EER	0.997	0.986	0.909	0.995	0.967	0.996	0.929
	IER	0.001	0.002	0.014	0.000	0.082	0.156	0.484

Table 4.3.2: Experimental and individual error rates for a situation with three equal-sized active effects (mean-shift of 2σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.663	0.336	0.022	0.522	0.295	0.673	0.000
	1	0.133	0.161	0.078	0.212	0.013	0.001	0.001
	2	0.103	0.165	0.223	0.146	0.080	0.004	0.015
	3	0.089	0.312	0.558	0.115	0.368	0.035	0.214
	4	0.009	0.022	0.103	0.004	0.084	0.024	0.241
	5	0.002	0.003	0.016	0.001	0.041	0.023	0.195
	6	0.000	0.001	0.002	0.000	0.028	0.023	0.148
	7	0.000	0.000	0.000	0.000	0.027	0.025	0.099
	8	0.000	0.000	0.000	0.000	0.021	0.025	0.050
	9	0.000	0.000	0.000	0.000	0.023	0.027	0.025
	10	0.000	0.000	0.000	0.000	0.022	0.024	0.010
	11	0.000	0.000	0.000	0.000	0.000	0.024	0.002
	12	0.000	0.000	0.000	0.000	0.000	0.027	0.000
	13	0.000	0.000	0.000	0.000	0.000	0.031	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.035	0.000
	EER	0.911	0.688	0.443	0.885	0.632	0.965	0.786
	IER	0.004	0.009	0.042	0.002	0.123	0.224	0.355

Table 4.3.3: Experimental and individual error rates for a situation with three equal-sized active effects (mean-shift of 4σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.224	0.027	0.000	0.122	0.023	0.478	0.000
	1	0.140	0.024	0.001	0.125	0.001	0.001	0.000
	2	0.170	0.047	0.015	0.188	0.013	0.004	0.000
	3	0.450	0.865	0.837	0.558	0.698	0.150	0.411
	4	0.013	0.032	0.123	0.007	0.096	0.043	0.284
	5	0.003	0.005	0.022	0.001	0.044	0.034	0.168
	6	0.001	0.001	0.003	0.000	0.033	0.035	0.086
	7	0.000	0.000	0.000	0.000	0.025	0.030	0.037
	8	0.000	0.000	0.000	0.000	0.023	0.029	0.012
	9	0.000	0.000	0.000	0.000	0.023	0.027	0.003
	10	0.000	0.000	0.000	0.000	0.022	0.032	0.000
	11	0.000	0.000	0.000	0.000	0.000	0.032	0.000
	12	0.000	0.000	0.000	0.000	0.000	0.034	0.000
	13	0.000	0.000	0.000	0.000	0.000	0.034	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.039	0.000
	EER	0.550	0.135	0.163	0.442	0.302	0.850	0.589
	IER	0.006	0.013	0.051	0.003	0.132	0.274	0.239

Table 4.3.4: Experimental and individual error rates for a situation with three equal-sized active effects (mean-shift of 6σ).

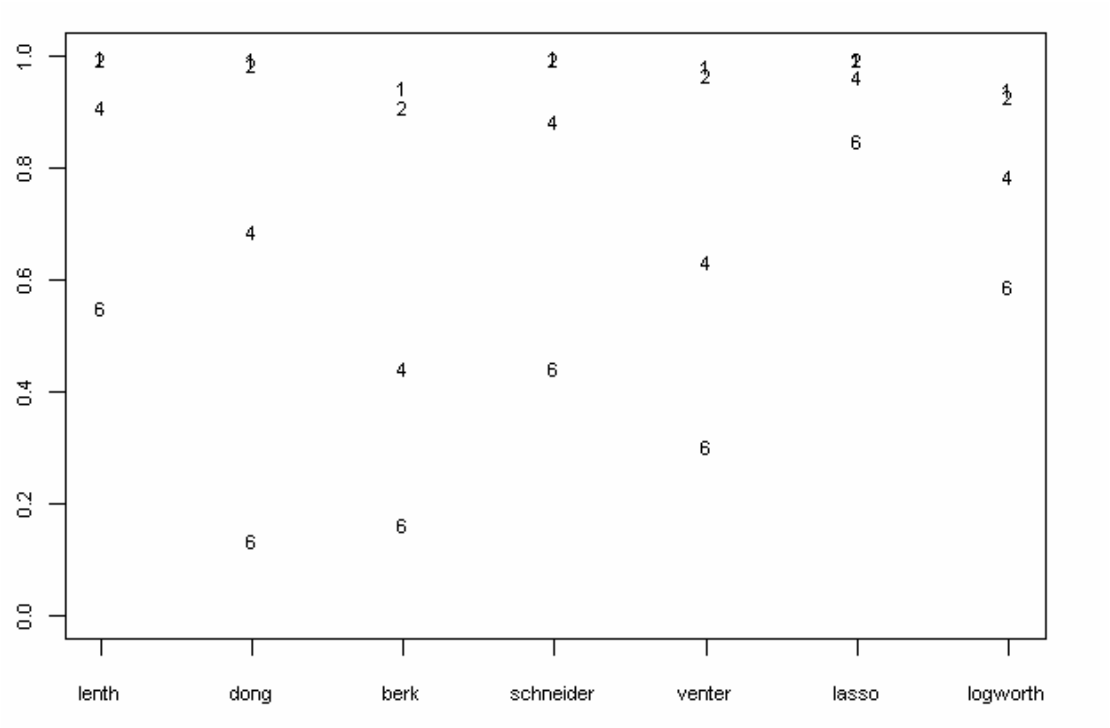

Figure 4.3.1: EER for situation with three equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

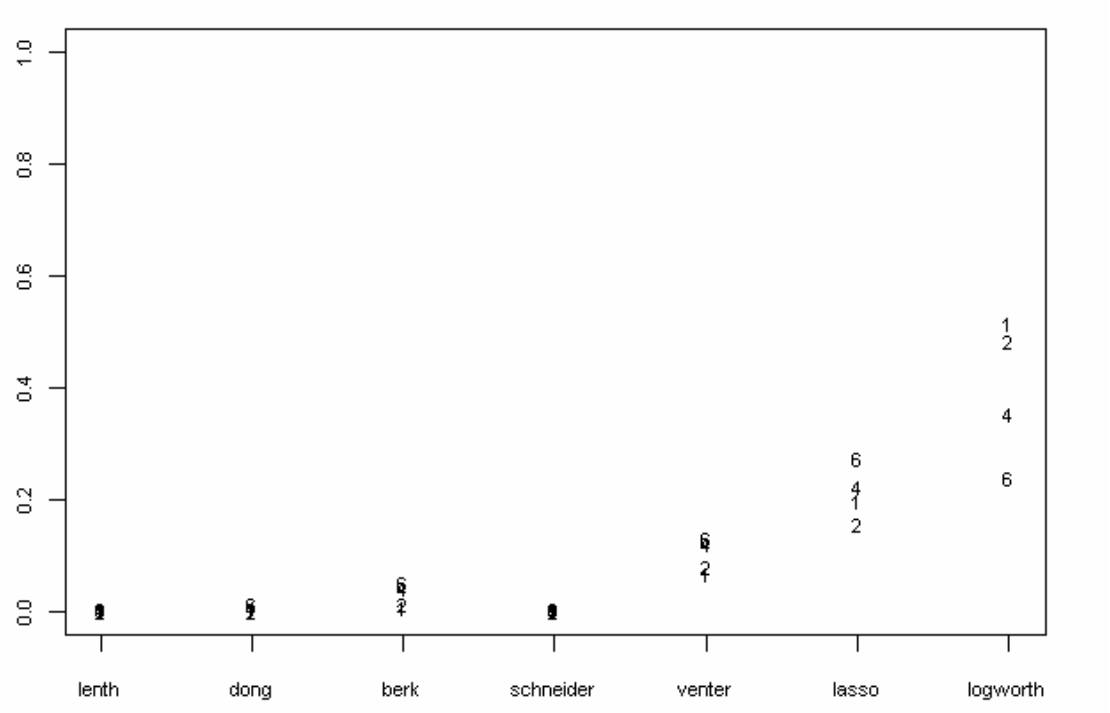

Figure 4.3.2: IER for situation with three equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

Large number of active effects

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.980	0.911	0.546	0.951	0.842	0.825	0.002
	1	0.014	0.070	0.263	0.041	0.015	0.000	0.035
	2	0.003	0.013	0.119	0.006	0.018	0.001	0.039
	3	0.002	0.005	0.052	0.001	0.016	0.003	0.060
	4	0.001	0.001	0.015	0.001	0.012	0.004	0.082
	5	0.001	0.001	0.004	0.000	0.013	0.007	0.114
	6	0.000	0.000	0.000	0.000	0.014	0.010	0.135
	7	0.000	0.000	0.000	0.000	0.017	0.013	0.153
	8	0.000	0.000	0.000	0.000	0.017	0.014	0.143
	9	0.000	0.000	0.000	0.000	0.018	0.014	0.113
	10	0.000	0.000	0.000	0.000	0.018	0.015	0.074
	11	0.000	0.000	0.000	0.000	0.000	0.018	0.036
	12	0.000	0.000	0.000	0.000	0.000	0.021	0.012
	13	0.000	0.000	0.000	0.000	0.000	0.023	0.002
	14	0.000	0.000	0.000	0.000	0.000	0.034	0.001
	EER	1.000	1.000	1.000	1.000	0.986	0.990	0.865
	IER	0.000	0.000	0.000	0.000	0.066	0.189	0.505

Table 4.4.1: Experimental and individual error rates for a situation with six equal-sized active effects (mean-shift of 1σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.977	0.933	0.513	0.957	0.819	0.907	0.001
	1	0.012	0.045	0.255	0.033	0.007	0.000	0.021
	2	0.006	0.012	0.130	0.007	0.011	0.000	0.024
	3	0.003	0.004	0.066	0.002	0.014	0.001	0.045
	4	0.001	0.003	0.027	0.001	0.020	0.003	0.079
	5	0.001	0.002	0.009	0.000	0.022	0.005	0.114
	6	0.001	0.001	0.002	0.000	0.022	0.006	0.149
	7	0.000	0.000	0.000	0.000	0.022	0.009	0.169
	8	0.000	0.000	0.000	0.000	0.020	0.010	0.161
	9	0.000	0.000	0.000	0.000	0.021	0.010	0.126
	10	0.000	0.000	0.000	0.000	0.021	0.008	0.075
	11	0.000	0.000	0.000	0.000	0.000	0.008	0.028
	12	0.000	0.000	0.000	0.000	0.000	0.008	0.009
	13	0.000	0.000	0.000	0.000	0.000	0.012	0.001
	14	0.000	0.000	0.000	0.000	0.000	0.013	0.000
	EER	0.999	0.999	0.998	1.000	0.978	0.994	0.852
	IER	0.000	0.000	0.000	0.000	0.079	0.092	0.531

Table 4.4.2: Experimental and individual error rates for a situation with six equal-sized active effects (mean-shift of 2σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.844	0.841	0.173	0.858	0.390	0.837	0.000
	1	0.040	0.043	0.162	0.066	0.000	0.000	0.001
	2	0.025	0.024	0.150	0.034	0.001	0.000	0.000
	3	0.021	0.019	0.150	0.018	0.004	0.000	0.001
	4	0.021	0.015	0.137	0.012	0.017	0.002	0.005
	5	0.021	0.014	0.129	0.007	0.077	0.007	0.032
	6	0.026	0.042	0.100	0.006	0.306	0.035	0.299
	7	0.002	0.002	0.000	0.000	0.097	0.021	0.305
	8	0.000	0.000	0.000	0.000	0.046	0.018	0.201
	9	0.000	0.000	0.000	0.000	0.033	0.015	0.104
	10	0.000	0.000	0.000	0.000	0.030	0.014	0.040
	11	0.000	0.000	0.000	0.000	0.000	0.012	0.011
	12	0.000	0.000	0.000	0.000	0.000	0.014	0.001
	13	0.000	0.000	0.000	0.000	0.000	0.011	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.015	0.000
	EER	0.974	0.958	0.901	0.995	0.694	0.965	0.701
	IER	0.002	0.001	0.000	0.000	0.181	0.135	0.579

Table 4.4.3: Experimental and individual error rates for a situation with six equal-sized active effects (mean-shift of 4σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.413	0.431	0.008	0.541	0.043	0.695	0.000
	1	0.060	0.048	0.015	0.109	0.000	0.000	0.000
	2	0.055	0.033	0.024	0.079	0.000	0.000	0.000
	3	0.047	0.029	0.040	0.065	0.000	0.000	0.000
	4	0.061	0.022	0.073	0.060	0.001	0.001	0.000
	5	0.084	0.022	0.167	0.060	0.019	0.006	0.000
	6	0.274	0.414	0.674	0.086	0.725	0.125	0.508
	7	0.007	0.001	0.000	0.000	0.105	0.037	0.308
	8	0.000	0.000	0.000	0.000	0.049	0.024	0.133
	9	0.000	0.000	0.000	0.000	0.031	0.022	0.041
	10	0.000	0.000	0.000	0.000	0.027	0.019	0.010
	11	0.000	0.000	0.000	0.000	0.000	0.017	0.001
	12	0.000	0.000	0.000	0.000	0.000	0.016	0.000
	13	0.000	0.000	0.000	0.000	0.000	0.018	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.018	0.000
	EER	0.726	0.586	0.326	0.914	0.275	0.875	0.492
	IER	0.006	0.001	0.000	0.000	0.186	0.190	0.410

Table 4.4.4: Experimental and individual error rates for a situation with six equal-sized active effects (mean-shift of 6σ).

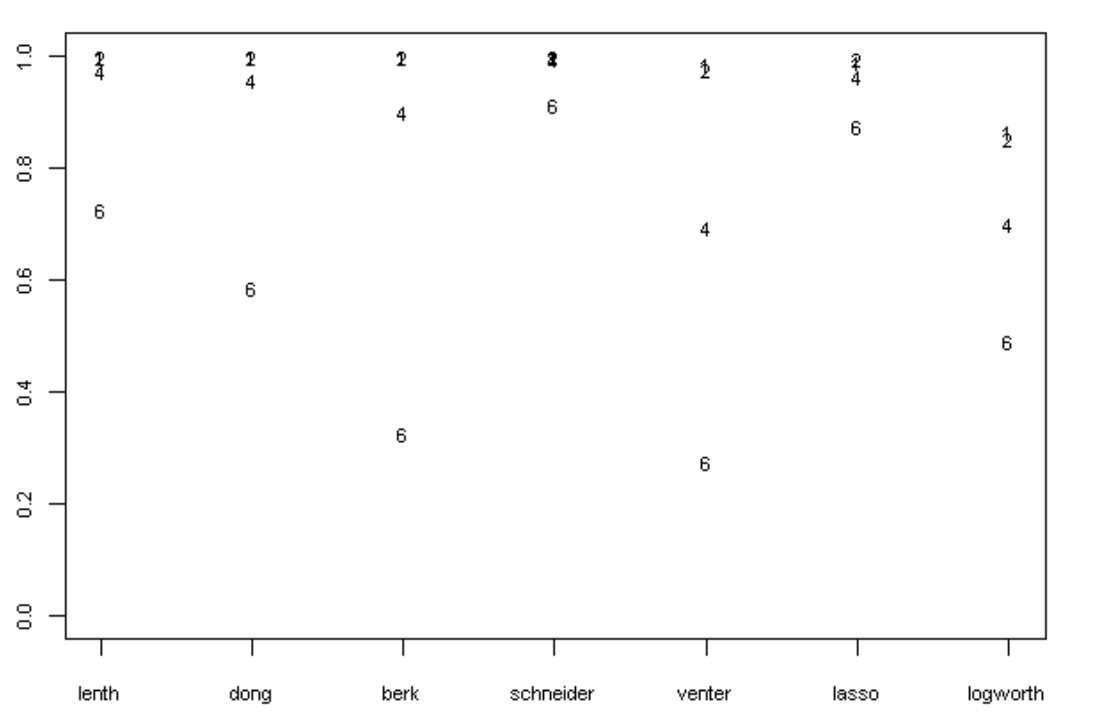

Figure 4.4.1: EER for situation with six equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

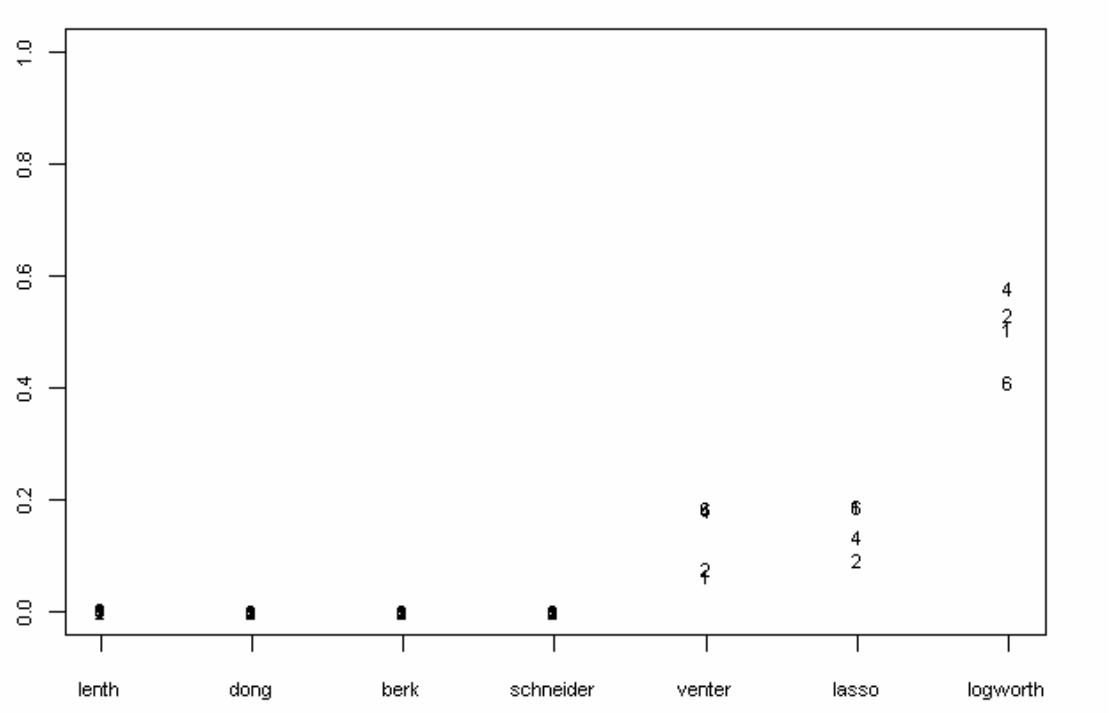

Figure 4.4.2: IER for situation with six equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.984	0.941	0.614	0.965	0.861	0.879	0.003
	1	0.010	0.043	0.223	0.028	0.010	0.000	0.029
	2	0.003	0.011	0.098	0.005	0.011	0.000	0.033
	3	0.003	0.003	0.044	0.002	0.012	0.001	0.051
	4	0.001	0.001	0.016	0.000	0.012	0.003	0.083
	5	0.000	0.001	0.004	0.000	0.013	0.005	0.105
	6	0.000	0.000	0.001	0.000	0.014	0.007	0.136
	7	0.000	0.000	0.000	0.000	0.015	0.009	0.150
	8	0.000	0.000	0.000	0.000	0.017	0.009	0.151
	9	0.000	0.000	0.000	0.000	0.018	0.012	0.118
	10	0.000	0.000	0.000	0.000	0.020	0.011	0.083
	11	0.000	0.000	0.000	0.000	0.000	0.011	0.042
	12	0.000	0.000	0.000	0.000	0.000	0.014	0.013
	13	0.000	0.000	0.000	0.000	0.000	0.016	0.003
	14	0.000	0.000	0.000	0.000	0.000	0.022	0.001
	EER	1.000	1.000	1.000	1.000	0.980	0.989	0.917
	IER	0.000	0.000	0.000	0.000	0.000	0.163	0.131

Table 4.5.1: Experimental and individual error rates for a situation with ten equal-sized active effects (mean-shift of 1σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.998	0.994	0.860	0.996	0.923	0.991	0.005
	1	0.001	0.005	0.100	0.003	0.001	0.000	0.025
	2	0.001	0.001	0.027	0.000	0.001	0.000	0.022
	3	0.000	0.000	0.008	0.000	0.002	0.000	0.033
	4	0.000	0.000	0.003	0.000	0.003	0.000	0.048
	5	0.000	0.000	0.001	0.000	0.004	0.000	0.078
	6	0.000	0.000	0.000	0.000	0.006	0.000	0.103
	7	0.000	0.000	0.000	0.000	0.010	0.001	0.135
	8	0.000	0.000	0.000	0.000	0.012	0.001	0.161
	9	0.000	0.000	0.000	0.000	0.017	0.001	0.158
	10	0.000	0.000	0.000	0.000	0.022	0.001	0.129
	11	0.000	0.000	0.000	0.000	0.000	0.001	0.070
	12	0.000	0.000	0.000	0.000	0.000	0.001	0.028
	13	0.000	0.000	0.000	0.000	0.000	0.002	0.006
	14	0.000	0.000	0.000	0.000	0.000	0.002	0.001
	EER	1.000	1.000	1.000	1.000	0.978	0.999	0.871
	IER	0.000	0.000	0.000	0.000	0.000	0.013	0.240

Table 4.5.2: Experimental and individual error rates for a situation with ten equal-sized active effects (mean-shift of 2σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	1.000	1.000	1.000	1.000	0.766	1.000	0.000
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.005
	2	0.000	0.000	0.000	0.000	0.000	0.000	0.004
	3	0.000	0.000	0.000	0.000	0.000	0.000	0.004
	4	0.000	0.000	0.000	0.000	0.000	0.000	0.004
	5	0.000	0.000	0.000	0.000	0.000	0.000	0.005
	6	0.000	0.000	0.000	0.000	0.000	0.000	0.008
	7	0.000	0.000	0.000	0.000	0.001	0.000	0.015
	8	0.000	0.000	0.000	0.000	0.006	0.000	0.027
	9	0.000	0.000	0.000	0.000	0.027	0.000	0.094
	10	0.000	0.000	0.000	0.000	0.201	0.000	0.463
	11	0.000	0.000	0.000	0.000	0.000	0.000	0.272
	12	0.000	0.000	0.000	0.000	0.000	0.000	0.087
	13	0.000	0.000	0.000	0.000	0.000	0.000	0.013
	14	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	EER	1.000	1.000	1.000	1.000	0.799	1.000	0.537
	IER	0.000	0.000	0.000	0.000	0.000	0.000	0.841

Table 4.5.3: Experimental and individual error rates for a situation with ten equal-sized active effects (mean-shift of 4σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	1.000	1.000	1.000	1.000	0.310	1.000	0.000
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	6	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	7	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	8	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	9	0.000	0.000	0.000	0.000	0.010	0.000	0.001
	10	0.000	0.000	0.000	0.000	0.680	0.000	0.683
	11	0.000	0.000	0.000	0.000	0.000	0.000	0.259
	12	0.000	0.000	0.000	0.000	0.000	0.000	0.051
	13	0.000	0.000	0.000	0.000	0.000	0.000	0.005
	14	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	EER	1.000	1.000	1.000	1.000	0.320	1.000	0.318
	IER	0.000	0.000	0.000	0.000	0.000	0.000	0.705

Table 4.5.4: Experimental and individual error rates for a situation with ten equal-sized active effects (mean-shift of 6σ).

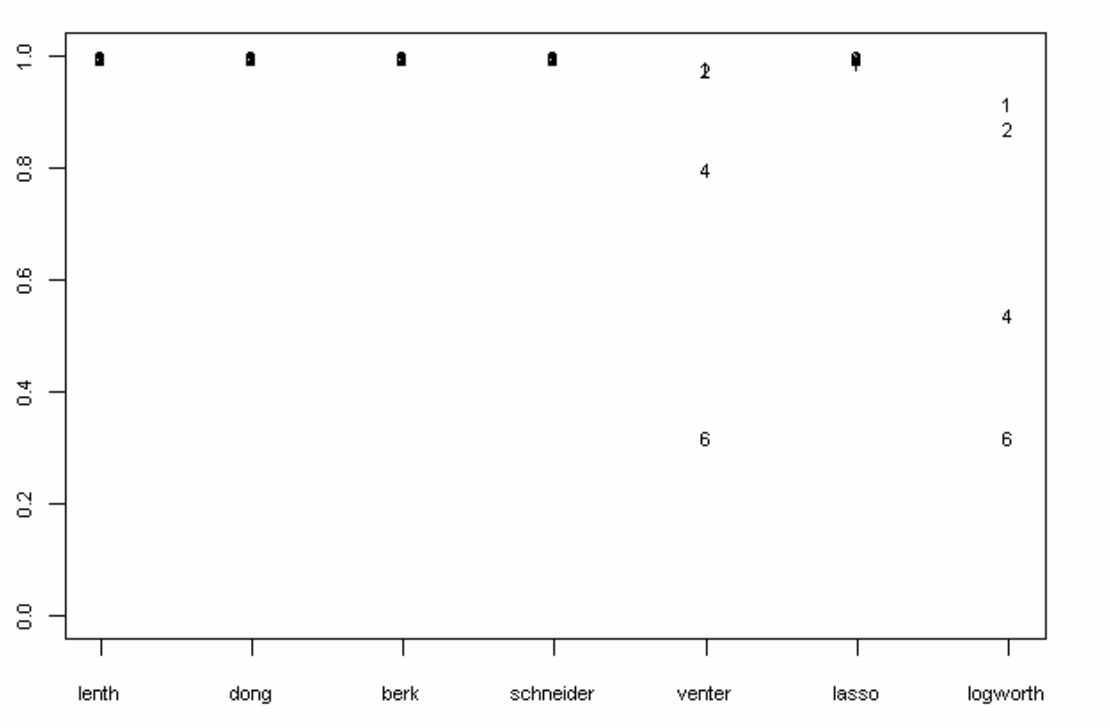

Figure 4.5.1: EER for situation with ten equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

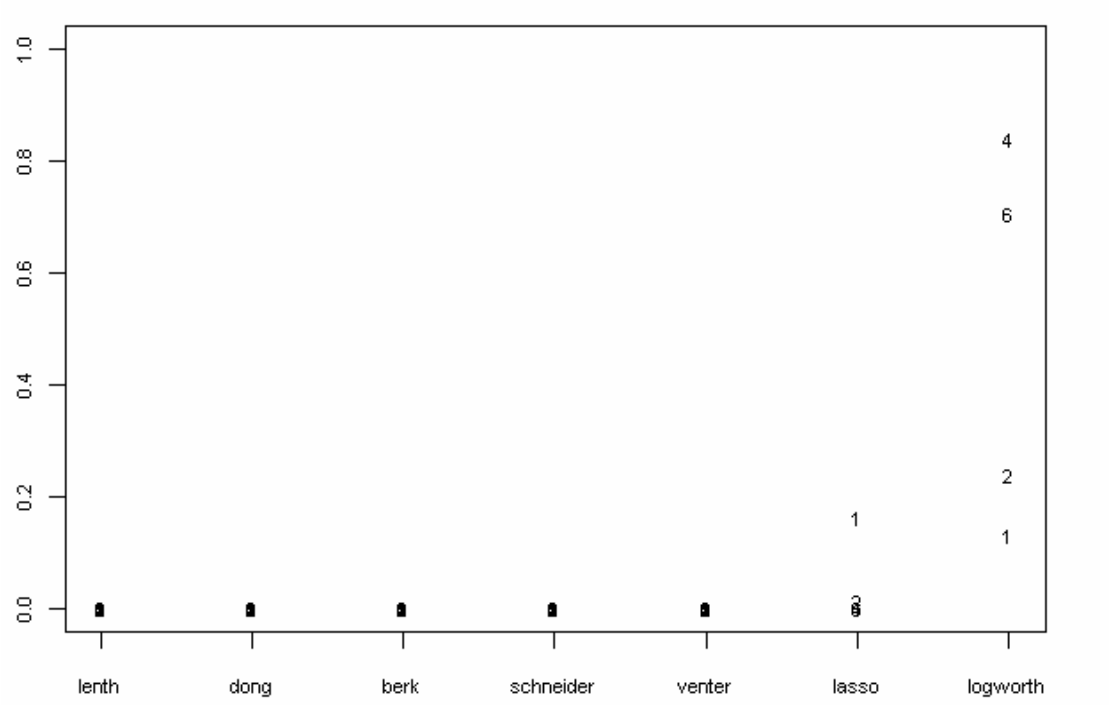

Figure 4.5.2: IER for situation with ten equal-sized active effects. The numbers (1, 2, 4, 6) indicate the active effect mean-shift in number of standard deviations.

Small number of large effects

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.010	0.000	0.000	0.000	0.001	0.290	0.000
	1	0.971	0.931	0.653	0.972	0.783	0.134	0.634
	2	0.014	0.055	0.229	0.024	0.050	0.034	0.214
	3	0.003	0.011	0.086	0.004	0.031	0.033	0.090
	4	0.001	0.002	0.025	0.000	0.022	0.031	0.040
	5	0.001	0.001	0.006	0.000	0.020	0.035	0.014
	6	0.000	0.000	0.001	0.000	0.017	0.035	0.006
	7	0.000	0.000	0.000	0.000	0.018	0.041	0.002
	8	0.000	0.000	0.000	0.000	0.019	0.039	0.000
	9	0.000	0.000	0.000	0.000	0.018	0.041	0.000
	10	0.000	0.000	0.000	0.000	0.021	0.045	0.000
	11	0.000	0.000	0.000	0.000	0.000	0.052	0.000
	12	0.000	0.000	0.000	0.000	0.000	0.060	0.000
	13	0.000	0.000	0.000	0.000	0.000	0.059	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.071	0.000
	EER	0.029	0.069	0.347	0.028	0.217	0.866	0.366
	IER	0.003	0.011	0.061	0.004	0.081	0.368	0.070

Table 4.6.1: Experimental and individual error rates for a situation with a single large active effect (mean-shift of 10σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.000	0.000	0.000	0.000	0.000	0.122	0.000
	1	0.981	0.934	0.654	0.971	0.783	0.359	0.762
	2	0.013	0.051	0.228	0.024	0.051	0.044	0.173
	3	0.004	0.010	0.085	0.005	0.028	0.037	0.049
	4	0.002	0.003	0.025	0.001	0.022	0.036	0.012
	5	0.001	0.001	0.008	0.000	0.020	0.037	0.003
	6	0.001	0.001	0.001	0.000	0.020	0.036	0.000
	7	0.000	0.000	0.000	0.000	0.018	0.035	0.000
	8	0.000	0.000	0.000	0.000	0.018	0.035	0.000
	9	0.000	0.000	0.000	0.000	0.020	0.035	0.000
	10	0.000	0.000	0.000	0.000	0.021	0.043	0.000
	11	0.000	0.000	0.000	0.000	0.000	0.048	0.000
	12	0.000	0.000	0.000	0.000	0.000	0.048	0.000
	13	0.000	0.000	0.000	0.000	0.000	0.045	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.042	0.000
	EER	0.019	0.066	0.346	0.030	0.217	0.641	0.238
	IER	0.004	0.011	0.061	0.005	0.082	0.305	0.040

Table 4.6.2: Experimental and individual error rates for a situation with a single large active effect (mean-shift of 15σ).

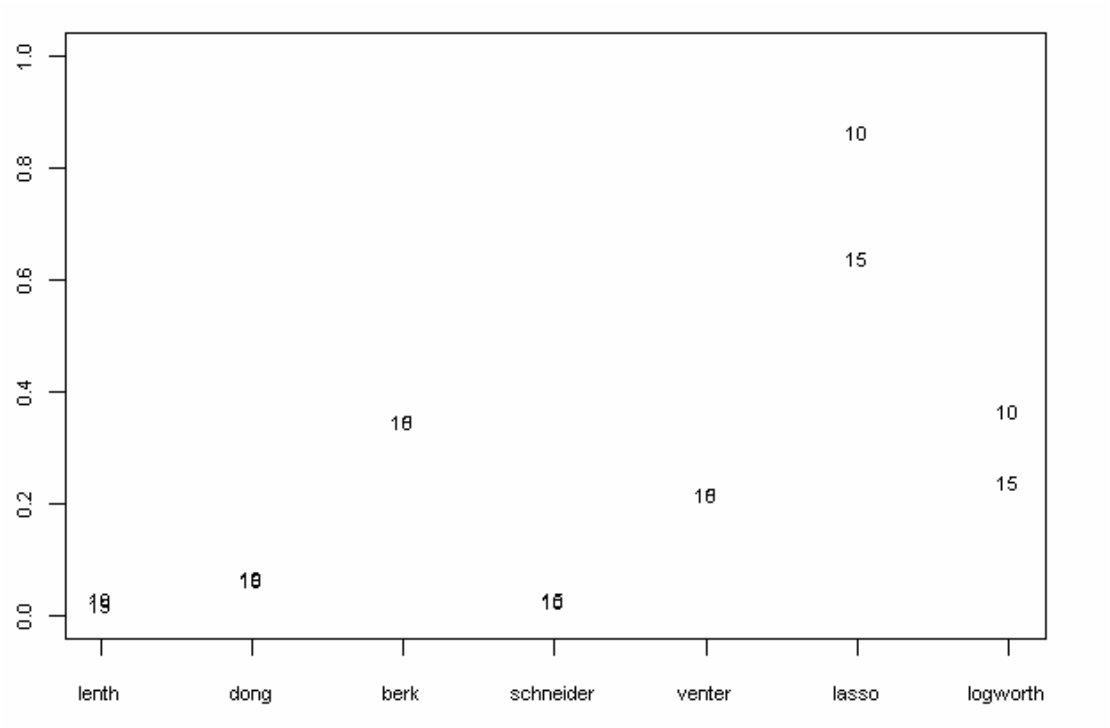

Figure 4.6.1: EER for situation with a single large active effect present. The numbers (10, 15) indicate the active effect mean-shift in number of standard deviations.

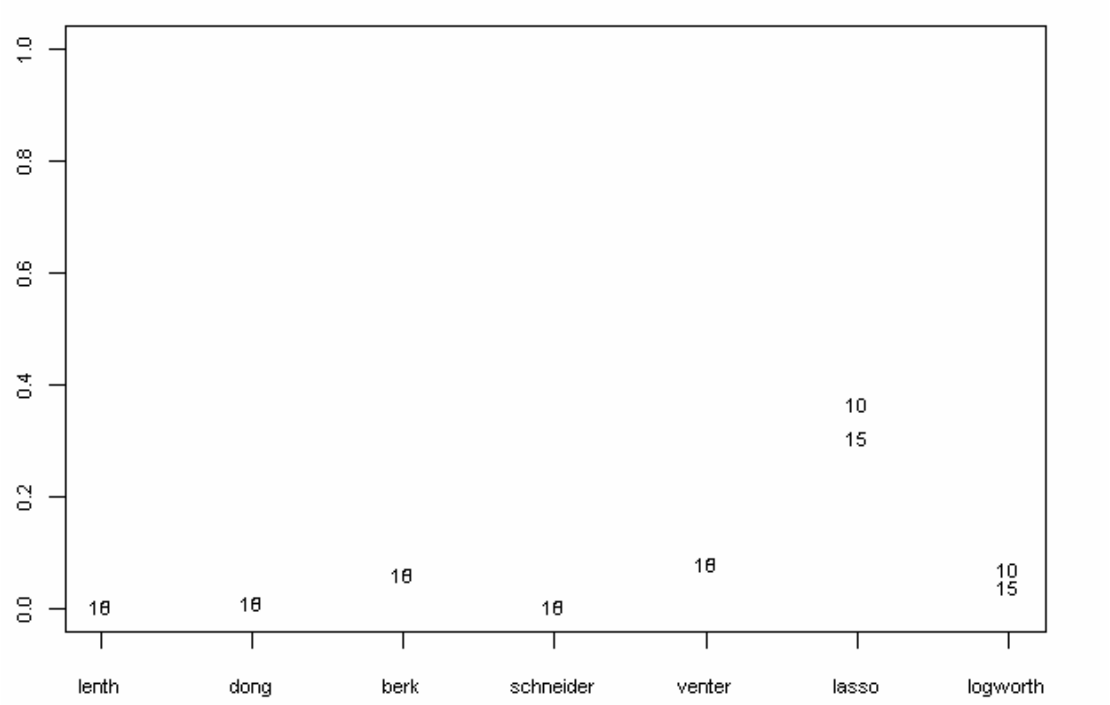

Figure 4.6.2: IER for situation with a single large active effect present. The numbers (10, 15) indicate the active effect mean-shift in number of standard deviations.

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.005	0.000	0.000	0.001	0.000	0.197	0.000
	1	0.006	0.000	0.000	0.001	0.000	0.001	0.000
	2	0.016	0.000	0.000	0.006	0.000	0.006	0.000
	3	0.957	0.966	0.863	0.985	0.740	0.459	0.636
	4	0.011	0.029	0.116	0.007	0.100	0.048	0.255
	5	0.004	0.004	0.018	0.001	0.040	0.034	0.079
	6	0.001	0.001	0.004	0.000	0.032	0.030	0.024
	7	0.000	0.000	0.000	0.000	0.025	0.028	0.006
	8	0.000	0.000	0.000	0.000	0.022	0.026	0.001
	9	0.000	0.000	0.000	0.000	0.020	0.026	0.000
	10	0.000	0.000	0.000	0.000	0.022	0.024	0.000
	11	0.000	0.000	0.000	0.000	0.000	0.029	0.000
	12	0.000	0.000	0.000	0.000	0.000	0.029	0.000
	13	0.000	0.000	0.000	0.000	0.000	0.032	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.032	0.000
	EER	0.043	0.034	0.138	0.015	0.260	0.541	0.364
	IER	0.006	0.012	0.048	0.003	0.128	0.246	0.133

Table 4.7.1: Experimental and individual error rates for a situation with three equal-sized large active effects (mean-shift of 10σ).

		Method						
		lenth	**dong**	**berk**	**schneider**	**venter**	**lasso**	**logworth**
Number of effects declared active	0	0.000	0.000	0.000	0.000	0.000	0.045	0.000
	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.000	0.000	0.000	0.000	0.000	0.004	0.000
	3	0.987	0.968	0.863	0.993	0.742	0.793	0.764
	4	0.011	0.027	0.115	0.007	0.096	0.025	0.185
	5	0.002	0.004	0.019	0.001	0.049	0.016	0.043
	6	0.001	0.001	0.002	0.000	0.032	0.015	0.007
	7	0.000	0.000	0.000	0.000	0.021	0.012	0.001
	8	0.000	0.000	0.000	0.000	0.020	0.014	0.000
	9	0.000	0.000	0.000	0.000	0.020	0.013	0.000
	10	0.000	0.000	0.000	0.000	0.021	0.015	0.000
	11	0.000	0.000	0.000	0.000	0.000	0.012	0.000
	12	0.000	0.000	0.000	0.000	0.000	0.014	0.000
	13	0.000	0.000	0.000	0.000	0.000	0.011	0.000
	14	0.000	0.000	0.000	0.000	0.000	0.013	0.000
	EER	0.013	0.032	0.137	0.007	0.258	0.207	0.236
	IER	0.005	0.011	0.048	0.002	0.126	0.112	0.084

Table 4.7.2: Experimental and individual error rates for a situation with three equal-sized large active effects (mean-shift of 15σ).

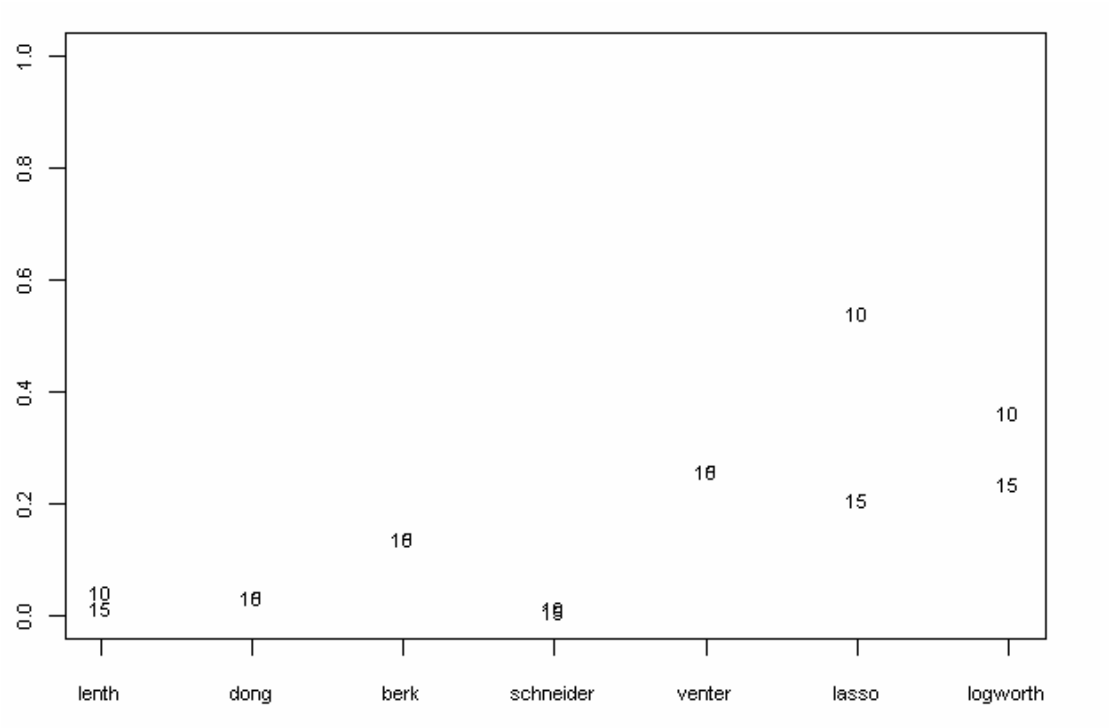

Figure 4.7.1: EER for situation with three equally sized large active effects. The numbers (10, 15) indicate the active effect mean-shift in number of standard deviations.

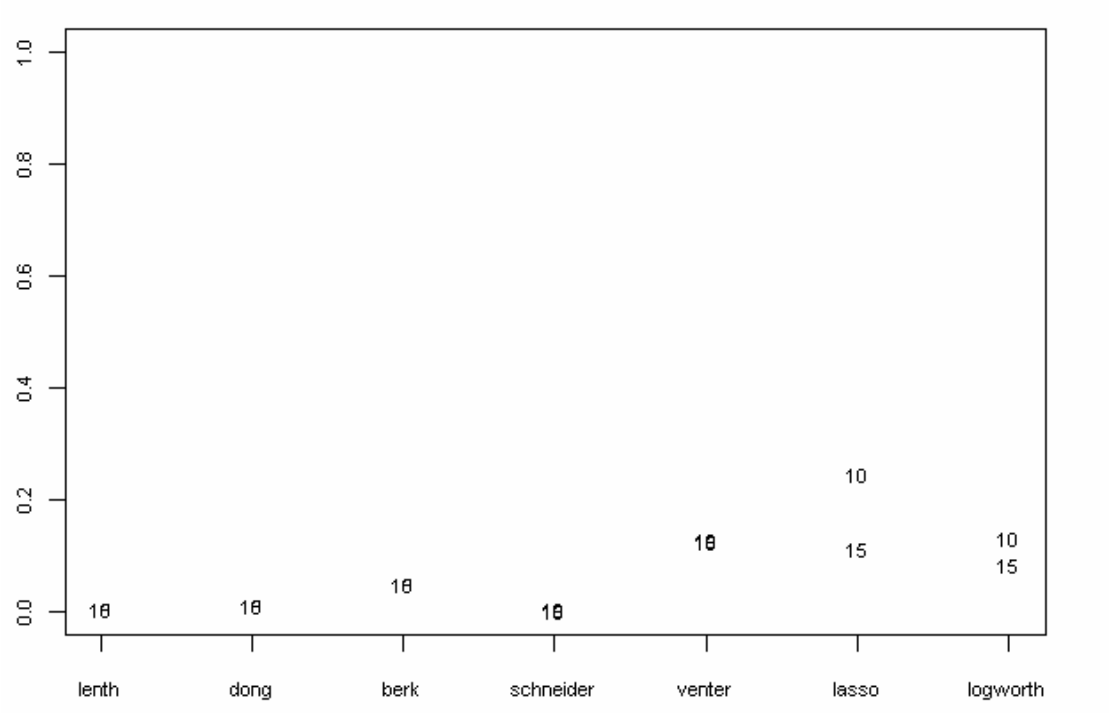

Figure 4.7.2: IER for situation with three equally sized large active effects. The numbers (10, 15) indicate the active effect mean-shift in number of standard deviations.

BIBLIOGRAPHY

A Simple Explanation of the Lasso and Least Angle Regression. Online: http://www-stat.stanford.edu/~tibs/lasso/simple.html. [Downloaded on October 31, 2006].

BERK, K.N., PICARD, R.R. 1991. *Significance Tests for Saturated Orthogonal Arrays* in Journal of Quality Technology, Vol. 23 No. 2. April 1991. pp 79-89.

BOX, G.E.P., DRAPER, N.R. 1987. *Empirical Model-Building and Response Surfaces.* John Wiley & Sons, Inc. USA. pp 47-50.

COCHRAN, W.G., COX, G.M. 1957. *Experimental Designs*. John Wiley & Sons, Inc. New York. pp 150-153.

CHRISTENSEN, R. 1996. *Analysis of Variance, Design and Regression: Applied statistical methods.* Chapman & Hall. Suffolk, Great Britain. p 178.

EFRON, B., HASTIE, T., JOHNSTONE, I., TIBSHIRANI, R. 2004. *Least Angle Regression* in Annals of Statistics, Vol. 32 No. 2. pp 407-499.

HAMADA, M., BALAKRISHNAN, N. 1998. *Analyzing Unreplicated Factorial Experiments: A Review With Some New Proposals* in Statistica Sinica, Vol. 8. pp 1-41.

HASTIE, T., EFRON, B. 2004. *lars: Least Angle Regression, Lasso and Forward Stagewise*. R package version 0.9-5. http://www-stat.stanford.edu/~hastie/Papers/#LARS

HINKLEMANN, K., KEMPTHORNE, O. 1994. *Design and Analysis of Experiments: Volume I: Introduction to Experimental Design*. John Wiley & Sons. USA. pp 352-359.

LENTH, R.V. 1989. *Quick and Easy Analysis of Unreplicated Factorials* in Technometrics, Vol. 31 No.4. November 1989. pp 469-473.

LOKHORST, J., VENABLES, B., TURLACH, B.; port to R and tests etc: MAECHLER, M. 2006. *lasso2: L1 constrained estimation aka `lasso'*. R package version 1.2-5. http://www.maths.uwa.edu.au/~berwin/software/lasso.html

MILLIKEN, G.A., JOHNSON, D.E. 1989. *Analysis of Messy Data – Volume 2: Nonreplicated Experiments.* Van Nostrand Reinhold. New York. pp 92-108.

MONTGOMERY, D.C. 2001. *Design and Analysis of Experiments*. John Wiley & Sons. USA. pp 170-276.

R DEVELOPMENT CORE TEAM. 2005. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

SAS INSTITUTE INC. 2003. *ADX: Automatic Selection Methods* in SAS Version 9.1 Help Files. SAS Institute Inc., Cary, NC, USA.

SCHNEIDER, H., KASPERSKI, W.J., WEISSFELD, L. 1993. *Finding Significant Effects for Unreplicated Fractional Factorials Using the* n *Smallest Contrasts* in Journal of Quality Tecnology, Vol. 25 No. 1. January 1993. pp 18-27.

SEARLE, S.R. 1971. *Linear Models.* John Wiley & Sons, Inc. USA. pp 75-224.

TIBSHIRANI, R. 1996. *Regression Shrinkage and Selection via the Lasso* in Journal of the Royal Statistical Society (Series B), Vol. 58 No. 1. pp 267-288.

VAN DER MERWE, A.S. 2004. *Variable Selection and Identification of Influential Observations in Multiple Linear Regression.* University of Stellenbosch: MComm thesis. Chapter 9 p 5. April 2004.

VENTER, J.H., STEEL, S.J. 1996. *A Hypothesis-Testing Approach Toward Identifying Active Contrasts* in Technometrics, Vol. 38 No. 2. May 1996. pp 161-169.

VINING, G. 1998. *Statistical Methods for Engineers*. Brooks/Cole Publishing Company. USA. pp 1-26, 353-450.