Variance estimation and multiple inference testing in saturated orthogonal designs

Tobias Adolfsson
Göteborg University and Chalmers University of Technology

30th April 2002

Abstract

This article concerns variance estimation and a step down multiple inference testing procedure in the case of orthogonal designs. The testing procedure is compared with other testing procedures and simulations show a considerable increase of power compared with other methods.

1 Introduction

To do experiments within industrial or technical research is often very expensive. One way of maximising the information gained while minimising the cost is to use statistical Design Of Experiments (DOE). An often used tool is, for instance, two level orthogonal design. Since time and money often are limiting factors the designs used are often reduced and without replicates. This leads to that all the degrees of freedom in the experiment will be consumed by the effect estimation. If this is the case we have no degrees of freedom left for estimating the variance and hence any formal inference will have to be of different type than the normal.

Let us assume that we are to perform a 2^{8-4}_{IV} designed experiment. The factors involved are called A, B, C, D, E, F G, and H. The person performing the experiment thinks it to be plausible that three main effect might be present. Let's say that A, B, C are believed to be true effects. If this is true the only true effects that there can be are from the factors and the

interactions A, B, C, AB, AC, BC and ABC. If we have access to replicates we could now estimate the variance and make inference like tests or confidence intervals for the effects. But since we believe that only seven of the 15 effects should be non-zero we actually have access to at least eight observations that could be used to estimate the noise. When using designed experiments as a screening process for optimising a process or product the situation mentioned above is often the case. In this case we could actually do inference. One should mention that when working with designed experiments, things like confounding patters are very important but in this article the focus will be on analysis of orthogonal designs not on DOE.

When analysing a saturated orthogonal experiment there are a number of different methods. Various graphical methods are available like normal probability plots and Pareto charts see, for instance, Box et al. (1978) for an overview. There have been attempts to formalise the graphical procedures, see Daniel (1959, 1976). In Daniel (1959) a half-normal plot was used when the modulus was plotted. This work was continued by Zahn (1975) who studied the empirical properties and simulated bounds for a testing procedure. For an overview of different methods for effect testing in saturated orthogonal designs see Haaland and O'Connel (1995) or Berk and Picard (1991).

All of these methods are based on the normation of the estimated effects by some kind of variance estimate. In Voss (1999) a more parametrical approach is used to construct the generalised modulus ratio (GMR) test. The GMR test is proven to maintain the correct multiple level of significance. In the proof of this property Voss uses a stochastic ordering lemma "borrowed" from the literature on ranking and selection. Voss also shows that the same method of proof will apply to the method in Berk and Picard (1991). Confidence intervals are also constructed and are shown to maintain, at least, the stated level of confidence. In the case of no true effect the confidence intervals are exact. The GMR test uses a simple but rather crude estimate for the variance. In this article we will use a slightly more complicated variance estimation process under the assumption that the measurement errors are normal. We will show that the test procedure that arises from this estimation will also maintain the correct multiple level of significance. We will also note that the proof of maintaining the correct level of significance might be translated to other distributions than the normal distribution. The asymptotic behaviour of the variance estimate will also be examined.

The setting is that we have $Y_i \sim N(\mu_i, \sigma)$, $i = 1, \ldots, n$ such that the Y_i are independent. We then use $|Y_i| = X_i$ as our sample. Assume that $\mu_i = 0$

say at least r of the X_i:s. The idea is to use the smallest r observations from the sample, which we hope contains no effect, and use these to estimate the variance. We will look at the r smallest observations of the sample and view these as a censured sample from a normal distribution and simply use the ml-estimates for σ as our estimate. The notation is not limited to saturated designs since the method might be applied to other problems.

2 Variance estimation

We will examine two different ways of estimating the variance in saturated designs. The methods used are actually more general than the orthogonal design setting, see Adolfsson (2002). In order to choose between the different estimates we will calculate the asymptotic variance of the estimates i.e. the Cramér-Rao limit. Both of the estimation methods have their positive and negative features. The first method views the observations as a censored sample which means that we take both the total number of effects and the number of effects used for estimating the variance into account. In the second method we view the sample as a sample from a cut off half-normal distribution. We then need to estimate both the cut off parameter and the variance. The support of the distribution is then dependent on the cut off parameter which makes asymptotic variance calculations difficult. The second method is however actually more correct in the setting when there exists effects since it really does not care what observations it misses. The true distribution to be used is more complicated since we in the first test have the situation used by the first method. If this test rejects we do have a sample from the r smallest out of n where we have one randomly selected observation has been excluded etc. This distribution changes in every step and hence so will the variance estimate. The use of such an estimate for use in a step down test procedure with correct multiple level of significance is, to my knowledge, an open problem.

2.1 Variance estimation method 1

As stated previously our sample consists of $X_j = |Y_j|$ where $Y_i \sim N(0, \sigma^2)$ $j = 1, \ldots, n$. We now order these observations into $0 < X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ and for notational convenience set $X_i = X_{(i)}$. Now we take the r smallest observations and use these for estimating σ in the ml-sense. The

density function for these censured data is

$$\binom{n}{r}\left(\prod_{i=1}^{r}\frac{2}{\sigma}\varphi\left(\frac{x_i}{\sigma}\right)\right)\left(2-2\Phi\left(\frac{x_r}{\sigma}\right)\right)^{n-r}$$

if x_r is the largest of the r smallest observations. This pdf is also the likelihood function for σ. So the log likelihood function is

$$l(\sigma,x)=k(n,r)-r\ln(\sigma)+(n-r)\ln(1-\Phi\left(\frac{x_r}{\sigma}\right))-\frac{1}{2\sigma^2}\sum_{i=1}^{r}x_i^2. \qquad (1)$$

Using that

$$\frac{\partial}{\partial\sigma}\varphi\left(\frac{x_i}{\sigma}\right)=\frac{x_i^2}{\sigma^3}\varphi\left(\frac{x_i}{\sigma}\right) \text{ and } \frac{\partial}{\partial\sigma}\Phi\left(\frac{x_r}{\sigma}\right)=-\frac{x_r}{\sigma^2}\varphi\left(\frac{x_r}{\sigma}\right)$$

we get

$$\frac{\partial l}{\partial\sigma}=0\Rightarrow-\frac{r}{\sigma}+(n-r)\frac{\frac{x_r}{\sigma^2}\varphi\left(\frac{x_r}{\sigma}\right)}{1-\Phi\left(\frac{x_r}{\sigma}\right)}+\frac{1}{\sigma^3}\sum_{i=1}^{r}x_i^2=0. \qquad (2)$$

A more convenient way of writing this equation is (both from a theoretical and practical point of view)

$$\sigma^2=\frac{r}{n-r}(\sigma^2-S_r^2)\frac{1-\Phi\left(\frac{x_r}{\sigma}\right)}{\frac{x_r}{\sigma}\varphi\left(\frac{x_r}{\sigma}\right)}. \qquad (3)$$

To get the ml-estimator we now have to solve this nonlinear equation. For comments on numerical methods for solving this see section 2.4.

One question that has to be answered is how large r has to be in relation to n for the estimate to be good enough. One way of getting a glimpse at this is to look at the asymptotic variance of the ml-estimate. It is known that under some regularity conditions a ml-estimator is asymptotically efficient (i.e asymptotically achieves the Cramér-Rao lower bound). The Cramér-Rao bound for the variance for the parameter $\tau(\theta)$ is given by

$$\frac{(\tau'(\theta))^2}{n\mathbb{E}_\theta\left[(\frac{\partial}{\partial\theta}\ln(f(X|\theta)))^2\right]}.$$

Since in our case $\tau(\sigma) = \sigma$ we only have to work with the denominator. We approximate the alternative expression for the denominator with

$$\begin{aligned} -\frac{\partial^2 l}{\partial \sigma^2} &= \frac{1}{\sigma^2}\left(3\sum_{i=1}^{r}\frac{x_i^2}{\sigma^2} + (n-r)\frac{\frac{x_r}{\sigma}\varphi\left(\frac{x_r}{\sigma}\right)}{1-\Phi\left(\frac{x_r}{\sigma}\right)}\right. \\ &\cdot \left.\left(2 - \frac{x_r^2}{\sigma^2} + \frac{\frac{x_r}{\sigma}\varphi\left(\frac{x_r}{\sigma}\right)}{1-\Phi\left(\frac{x_r}{\sigma}\right)}\right) - r\right). \end{aligned} \tag{4}$$

If we let $r/n = p \in (0,1)$ and $r, n \to \infty$ we get

$$u_r = \frac{x_r}{\sigma} \to u_\infty = \Phi^{-1}\left(\frac{p}{2} + \frac{1}{2}\right). \tag{5}$$

One can also see that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{r}\frac{X_i^2}{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{r}U_i^2 = \frac{1}{n}\sum_{i=1}^{n}U_i\mathbb{I}_{\{U_i \le U_r\}} \\ &\to \mathbb{E}[U^2\,|U < u_\infty] = \int_0^{u_\infty} u^2 2\varphi(u)\,du \end{aligned} \tag{6}$$

using the law of large numbers. Using (5) and (6) we get an approximation of (4) valid for large n and r, namely

$$\begin{aligned} -\frac{\partial^2 l}{\partial \sigma^2} &\approx \frac{n}{\sigma^2}\left(2u_\infty\varphi(u_\infty)\left(2 + 2\frac{u_\infty\varphi(u_\infty)}{1-p} - u_\infty^2\right)\right. \\ &+ \left. 3\int_0^{u_\infty} u^2 2\varphi(u)\,du - p\right) \\ &= \frac{n}{\sigma^2}f(p). \end{aligned}$$

The asymptotic factor function $f(p)$ is plotted in figure 1. This can be compared to the variance of the variance estimate that one would derive if one used the complete set of data. Calculating this is a simple task and one gets

$$-\frac{\partial^2 l_1}{\partial \sigma^2} \approx \frac{n}{\sigma^2}\left(3\mathbb{E}[U^2] - 1\right) = \frac{2n}{\sigma^2}.$$

Some elementary calculus gives that the

$$2u_\infty\varphi(u_\infty)\left(2 + 2\frac{u_\infty\varphi(u_\infty)}{1-p} - u_\infty^2\right)$$

part of $f(p)$ goes to zero as p approaches 1 from below, and hence the two asymptotic variances are equal at $p = 1$.

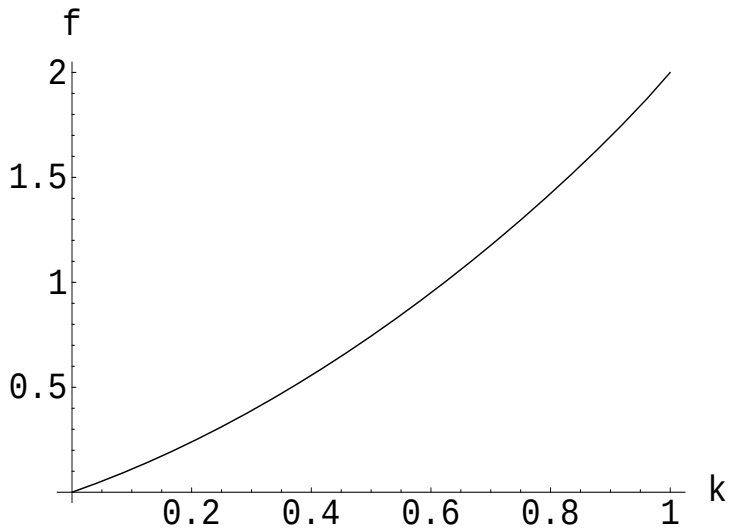

Figure 1: Asymptotic variance factor for the first method

2.2 Variance estimation method 2

The method presented here has the advantage of being more robust for the random censoring that will occur in the testing procedure that we are to construct. On the other hand it is more dependent on how large a proportion of the sample that we use for estimation since it uses less information than the previous method.

The idea now is to simply cut off the distribution at some level Δ. We then estimate σ (and Δ) in the maximum likelihood sense. So we now see the observations as observations from a distribution living only on $(0, \Delta)$ and hence the likelihood function is

$$L(\sigma, \Delta, x) = \prod_{i=1}^{r} \frac{\frac{2}{\sigma}\varphi\left(\frac{x_i}{\sigma}\right)}{2\left(\Phi\left(\frac{\Delta}{\sigma}\right) - 1/2\right)}, \tag{7}$$

and the log-likelihood function is then

$$l(\sigma, \Delta, x) = -n \ln \sigma + r - \frac{1}{2\sigma^2} \sum_{i=1}^{r} x_i^2 - n \ln \left(\Phi\left(\frac{\Delta}{\sigma}\right)\right) + k \tag{8}$$

where k is a known constant. From (7) we see that Δ should be estimated with the lowest possible value that makes sense, i.e $\Delta = x_r$. We put this into

(8), differentiate and get

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} - \frac{1}{2\sigma^2}\sum_{i=1}^{r} x_i^2 + n\frac{\frac{x_r}{\sigma^2}\varphi\left(\frac{x_r}{\sigma}\right)}{\Phi\left(\frac{x_r}{\sigma}\right) - 1/2}.$$

Setting this to zero and solving gives

$$\sigma^2 = \frac{(\sigma^2 - s_r^2)(\Phi\left(\frac{x_r}{\sigma}\right) - 1/2)}{\frac{x_k}{\sigma}\varphi\left(\frac{x_r}{\sigma}\right)}. \tag{9}$$

This is also a nonlinear equation that has to be solved numerically. In the case of this estimate the asymptotics are not as clear as in the previous case. The problem is that the set for which $f_X(x) \neq 0$ is dependent on the parameter Δ. This might result in that the Cramér-Rao bound might not be valid for σ and super efficient estimates might exist. However, if Δ and σ estimates are independent the Cramér-Rao bound will hold for σ. In the present case we can not at all be sure that the asymptotic results hold but never the less they will be presented, **keep this in mind!** Following the methodology in the previous section we calculate

$$\begin{aligned} -\frac{\partial^2 l}{\partial \sigma^2} &= \frac{n}{\sigma^2}\left(1 - \frac{3}{n}\sum_{i=1}^{r}\frac{x_i^2}{\sigma^2} + \frac{x_r}{\sigma}\left(\frac{-2\varphi\left(\frac{x_r}{\sigma}\right) + \frac{x_r^2}{\sigma^2}\varphi\left(\frac{x_r}{\sigma}\right)}{\Phi\left(\frac{x_r}{\sigma}\right) - 1/2}\right.\right. \\ &+ \left.\left.\frac{x_r}{\sigma}\left(\frac{\varphi\left(\frac{x_r}{\sigma}\right)}{\Phi\left(\frac{x_r}{\sigma}\right) - 1/2}\right)^2\right)\right). \end{aligned}$$

Using the same notation and arguments as in the previous section we get

$$\begin{aligned} -\frac{\partial^2 l}{\partial \sigma^2} &\to \frac{n}{\sigma^2}\left(3\int_0^{u_\infty} u^2\varphi(u)\,du - p - u_\infty\varphi(u_\infty)\left(-4 + u_\infty^2\right.\right. \\ &+ \left.\left. 2\frac{u_\infty\varphi(u_\infty)}{p}\right)\right). \end{aligned}$$

Let us write this on the form

$$-\frac{\partial^2 l}{\partial \sigma^2} \to \frac{n}{\sigma^2} f(p).$$

This is valid since u_∞ is a function of p. We would like $f(p) \to 0$ as $p \to 0$ and $f(p) \to 2$ as $p \to 1$. The latter limit is trivial but for the first limit to

hold we need to see that

$$\lim_{p\to 0}\frac{u_\infty(p)}{p}=\lim_{p\to 0}\frac{u'_\infty(p)}{1}=\lim_{p\to 0}\frac{1}{p'(u_\infty)}=\lim_{p\to 0}\frac{1}{2\varphi\left(u_\infty\right)}=\frac{\sqrt{2\pi}}{2}=\sqrt{\frac{\pi}{2}}$$

using l'Hopital's rule and implicit derivatives. We can now conclude that the function f has the same behaviour in the endpoints as the function in the previous section. This function does however grow more rapidly close to 2 and is hence less effective in some sense. Figure 2 shows the plot of $f(p)$.

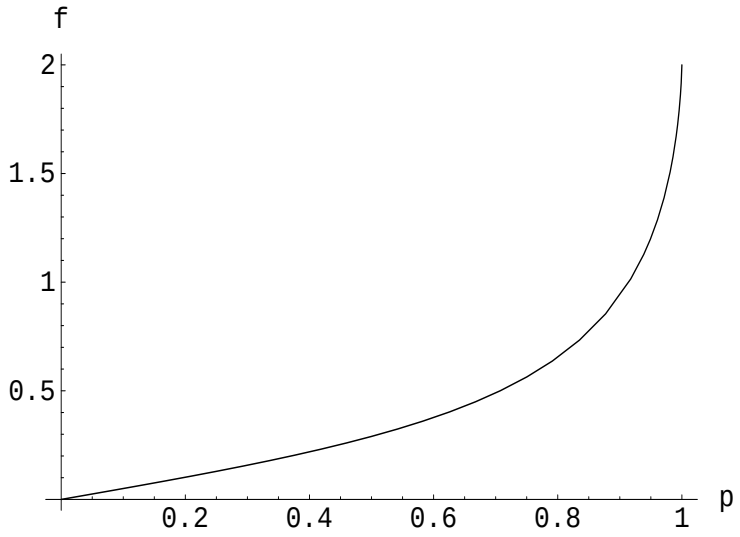

Figure 2: Asymptotic variance factor of the second method

2.3 Choosing variance estimation method

As we could see from the asymptotic variances, method one, viewing the problem as a censored sample, is the better of the two. This is natural since we actually have more information in this case. We have included in our model that we have n observations from the beginning. In the case of the second method we view the sample as from a cut off half-normal distribution. If the sample proportion used for estimation is small, which probably would be the case in saturated orthogonal designs, the cut off level will be low. This imply that the sample will come from a region of the normal

distribution close to the origin. Since we are trying to estimate the curvature of the normal distribution this puts us in a rather difficult position since the normal distribution is rather flat in this region. The flatness of this region makes the estimates unreliable and hence we choose to work with the first of the variance estimations techniques.

2.4 Numerical comments on variance estimation

The function (1) and equation (2) might look a bit complicated but they are quite easy to handle. The simplest method for assessing the ml-estimator is probably to solve (2). When this function is plotted it gives an L-shaped curve which is known to give a complicated equation to solve using normal Newton methods based on estimation of the derivatives. I propose that a simple interval splitting algorithm is used. One could use Möbius approximations and solve these instead but this seems unnecessary. One should check whether or not the algorithm one is about to use uses derivatives or not since methods using derivatives might give very poor convergence (or even no convergence). There is an important point to this: many estimations using the ML-method give rather difficult numerical problems and one should be careful and make sure that the optimising tool that is used is actually suitable for the task.

3 Testing for true effects in saturated orthogonal designs

We are going to construct a step down test for testing for true effects that maintains the correct multiple level of significance and different types of confidence intervals that both keeps the correct multiple level of significance and some that only keep the individual level. In the first section we will work out the theory needed for showing these results. Hence the underlying model has the structure (since we are dealing with normal distributions folded round the y-axis we can assume that $\mu \geq 0$)

$$\mu_{(1)} \geq \mu_{(2)} \geq \ldots \geq \mu_{(i)} \geq \mu_{(i+1)} = \ldots = \mu_{(n)} = 0.$$

The problem is now to find i such that $\mu_{(i)} > 0$ and $\mu_{(i+1)} = 0$.

3.1 Theory

To show that the test that we are about to construct gives the correct multiple level of significants we need a lemma. The main idea in the proof of correct multiple level of significance is the use of stochastic orderings. To show this monotonicity of the test statistic we need to examine the variance estimates further. In the proof of this lemma we also need a fact known from reliability theory namely that the failure rate of a normal distribution is monotonically increasing. An observation to be made is that the same arguments will do for other distributions in the exponential family with monotony increasing failure rates (see Adolfsson (2002) for the gamma distribution).

Lemma 1 *The estimate σ^2 given by (3) is non-decreasing in S_r^2 for fix x_r.*

Proof: We will use a contra positive proof. We will show that if σ^2 decreases so will S_r^2. We rewrite (3) in the form

$$\sigma^2 = \frac{r}{n-r}(\sigma^2 - S_r^2) f\left(\frac{x_r}{\sigma}\right).$$

Rearranging the equation gives

$$\sigma^2 \left(1 - \frac{n-r}{rf\left(\frac{x_r}{\sigma}\right)}\right) = S_r^2.$$

We observe that $f(x) = 1/(xz(x))$ where $z(x)$ is the failure rate function for a normal distribution. We know now that $f(x)$ is a decreasing function. The proof is now easy: σ^2 decreases $\Rightarrow$ x_r/σ increases $\Rightarrow$ $f(x_r/\sigma)$ decreases $\Rightarrow$ $\frac{n-r}{rf\left(\frac{x_r}{\sigma}\right)}$ increases $\Rightarrow$ $1 - \frac{n-r}{rf\left(\frac{x_r}{\sigma}\right)}$ decreases $\Rightarrow$ $\sigma^2 \left(1 - \frac{n-r}{rf\left(\frac{x_r}{\sigma}\right)}\right)$ decreases $\Rightarrow$ S_r^2 decreases. **Q.E.D.**

Lemma 2 *The estimate σ^2 given by (3) is non-decreasing in x_r.*

Proof: Once more we rearrange equation (3) to get

$$\frac{n-r}{rf\left(\frac{x_r}{\sigma}\right)} + \frac{1}{r} \cdot \frac{x_r^2}{\sigma^2} = 1 - \frac{k}{\sigma^2} \tag{10}$$

where k is a constant. We observe that σ is a non-constant function of x_r. Now take $x_r' \in \mathbb{R}^+$ such that $x_r < x_r'$. We observe that when x_r/σ

increases then (10) gives that σ increases. We are now going to force out a contradiction. Let $\sigma' = \sigma(x_r')$, assume that

$$\frac{x_r}{\sigma} > \frac{x_r'}{\sigma'}.$$

According to (10) we must have $\sigma' < \sigma$ but this gives that $x_r > x_r'$, a contradiction, and hence

$$\frac{x_r}{\sigma} \leq \frac{x_r'}{\sigma'}.$$

Once again we use (10) and conclude that $\sigma \leq \sigma'$. **Q.E.D.**

A basic part of the proof of correct multiple level of significants is stochastic ordering. The following definition and lemma is needed.

Definition 1 (Lehmann (1986)) *A family of distribution functions $F_\theta(x)$ on $\mathbb{R}$, for θ a real parameter, is said to be* stochastically increasing *if $\theta < \theta'$ implies $F_\theta(x) \geq F_{\theta'}(x)\ \forall x$. Similarly, the family of distributions is said to be* stochastically decreasing *if $\theta < \theta'$ implies $F_\theta(x) \leq F_{\theta'}(x)\ \forall x$.*

Lemma 3 (Stochastic ordering lemma (Voss (1999))) *Let $F_\theta(x)$, for real parameters θ, be a stochastically increasing family of distribution functions on the real line. Let $X_1, X_2, \ldots, X_k$ be independent random variables, where the distribution function of X_i is $F_{\theta_i}(x_i)$. For any fixed i $(1 \leq i \leq k)$, if the statistic $t = t(x_1, x_2, \ldots, x_k)$ is non-increasing (nondecreasing) function of x_i when all x_j for $j \neq i$ are held fixed, then the distribution of $T = t(X_1, X_2, \ldots, X_k)$ is stochastically decreasing (increasing) in θ_i.*

3.2 Test construction

We are going to construct a test based on the statistic

$$T_i = \frac{\max\{X_1, \ldots, X_i\}}{\hat{\sigma}}$$

which we are going to compare with critical values $c_\alpha(i, r, n)$ from the distribution of

$$C_i = \frac{\max\{X_1, \ldots, X_i\}}{\hat{\sigma}}, \quad i = 1, \ldots, n$$

such that

$$\mathbb{P}\left\{T_i > c_\alpha(i, r, n)\right\} = \alpha$$

under the null-hypothesis. The indexes r and n in $c_\alpha(i,r,n)$ is due to the fact that $\hat{\sigma} = \hat{\sigma}(r,n)$. With this we will be able to construct a testing method that maintains a correct multiple level of significant.

For this test to have any practical relevance we need to show that $\hat{\sigma}/\sigma$ is independent of σ. We note that the distribution of X_i/σ is independent of σ and using (2) we get that

$$\frac{\hat{\sigma}}{\sigma} = \frac{\frac{X_r}{\sigma}(p-r)\varphi\left(\frac{X_r}{\hat{\sigma}}\right)}{\left(1-\Phi\left(\frac{X_r}{\hat{\sigma}}\right)\right)\left(r - \frac{1}{\hat{\sigma^2}}\sum_{i=1}^{r} X_i^2\right)}.$$

Substituting $Z_i = X_i/\sigma$ and $q = \hat{\sigma}/\sigma$ we get

$$q = \frac{Z_r(p-r)\varphi\left(\frac{Z_r}{q}\right)}{\left(1-\Phi\left(\frac{Z_r}{q}\right)\right)(r - \frac{1}{q^2}\sum_{i=1}^{r} Z_i^2)}$$

and we see that this solution (in q) is independent of σ. This tells us that the test statistic T_i is independent of σ and hence that the test may be of practical use.

The test procedure is to reject the hypothesis $H_{0,n} : \mu_n = 0$ and conclude $H_{1,n} : \mu_n \neq 0$ if $t_n > c_\alpha(n,r,n)$ otherwise **no** rejections are made. If we reject in the first test we test $H_{0,n-1} : \mu_{n-1} = 0$ against $H_{1,n-1} : \mu_{n-1} \neq 0$ and reject if $t_{n-1} > c_\alpha(n-1,r,n)$. If no rejection is made in the second step we stop and only conclude that $\mu_n \neq 0$, otherwise we conclude that $\mu_n \neq 0$ and $\mu_{n-1} \neq 0$ and we continue. The procedure is now repeated until the first time we can not reject. Hence, no hypothesis can be rejected if not all of its predecessors have been rejected. One observation to be made is that the deeper we go down the rejection chain the more difficult it will be to reject. This is due to our variance estimate since if we have rejected a false hypothesis the variance estimation is incorrect. The correct variance estimate should be one where we have the r smallest out of n where one observation has been randomly excluded. We will over estimate the variance and hence our method will be conservative when true effects are present.

We now show that this test procedure maintains the correct level of significants.

Theorem 1 *If X_i $(i = 1,\ldots,n)$ are independent and $X_i = |Y_i|$, $Y_i \sim N(\mu_i,\sigma^2)$, then the test procedure above maintains the multiple level of significants α. The multiple significants level is exactly α in the null case.*

Proof: Let h denote the number of effects μ_i that are zero. The error rate is exactly α in the null case, for which $h = n$, from the definition of $c_\alpha(n, r, n)$. If $h = 0$ we can not do any false rejections. Hence we assume that $1 \leq h < n$. Without loss of generality we relabel the effects so that $\mu_i = 0$ for $i \leq h$ and $\mu_i \neq 0$ otherwise. A false assertion can only be made if

$$t_h = \frac{\max_{i \leq h}\{x_i\}}{\hat{\sigma}} > c_\alpha(h, r, n).$$

Now according to lemma 1 and lemma 2 t_h is a non-increasing function of x_j for each $j > h$, so the distribution of T_h is stochastically decreasing in each μ_j for each $j > h$ by the stochastic ordering lemma. We note that the critical values are chosen such that

$$\mathbb{P}\{T_h > c_\alpha(h, r, n)\} = \alpha$$

when $\mu_i = 0$ for all i. Hence,

$$\mathbb{P}\{T_h > c_\alpha(h, r, n)\} \leq \alpha$$

for **any** parameter configuration $(\mu_1, \mu_2, \ldots, \mu_n)$. **Q.E.D.**

n	k										
8	4	4.584	3.732	2.846	1.967						
9	4	4.789	4.160	3.307	2.549	1.747					
10	5	4.228	3.747	3.058	2.434	1.758					
11	5	4.456	3.956	3.359	2.755	2.168	1.567				
12	6	4.102	3.627	3.148	2.698	2.163	1.589				
13	6	4.238	3.817	3.334	2.956	2.432	1.979	1.477			
14	7	4.003	3.634	3.291	2.824	2.376	1.969	1.491			
15	7	4.184	3.772	3.407	3.080	2.583	2.172	1.848	1.384		
16	8	3.998	3.659	3.313	2.960	2.573	2.214	1.854	1.438		
17	8	4.040	3.742	3.406	3.048	2.709	2.442	2.061	1.737	1.355	
18	9	3.912	3.666	3.352	3.009	2.677	2.382	2.094	1.745	1.386	
19	9	3.953	3.701	3.426	3.106	2.834	2.557	2.247	1.975	1.649	1.316
20	10	3.906	3.629	3.358	3.045	2.814	2.554	2.236	1.988	1.687	1.351

Table 1: Table of critical values for 5% significance given n and k.

3.3 Simulating critical values for the testing procedure

The speed of our computers is still growing by the hour. This gives that simulations will be easier and easier to perform. But on the other hand we will also want to solve larger and larger problems. It is my experience that the implementation of the simulation method makes quite a difference. To generate the table 1, which contains critical values for the testing procedure in the previous section with an level of significance equal to 5% (based on 10000 simulations each), I used a specially designed C-program. Comparing the executional times between MATLAB and my code is comparing days to minutes. One should be aware that even though our computers are fast, and getting faster, we can still gain a lot from good algorithms.

4 Construction of confidence intervals

The construction of multiple confidence intervals are closely related to the method used for making intervals using the studentizied maximum modulus distribution. The intervals will only be calculated for effect not used in the variance estimate $\hat{\sigma}^2$. One could actually make intervals for all effects but then one needs to calculate a new distribution. We would have to construct a new distribution where we make sure that the X_i in the numerator is not used in the denomerator. By this we would avoid dependence between numerator and denomerator in the pivot distribution needed. Of course intervals constructed this way will be conservative.

The construction of a confidence interval that maintains the multiple level of significance is straight forward. We will use the same notation as in the previous section. The pivotal distribution will be that of

$$\max_i \frac{X_i}{\hat{\sigma}} \tag{11}$$

assuming that $\mu_i = 0$. The interval is derived from

$$1-\alpha \;=\; \mathbb{P}\left\{\max_i \frac{|Y_i-\mu_i|}{\hat{\sigma}} < z_\alpha\right\}. \tag{12}$$

Since this is valid for all μ_i:s the intervals will maintain the correct level of significance.

For constructing individual confidence intervals one needs to calculate a new pivotal distribution. To make this we need a new notation σ_i. The distribution needed is that of

$$\frac{|Y_i - \mu_i|}{\hat{\sigma}_i} \tag{13}$$

where $\hat{\sigma}_i$ is the estimate of σ using the r smallest observations *excluding* X_i. We need to do this to be able to use the stochastic ordering lemma to show that this distribution is stochastically decreasing in all $\mu_j,\ j \neq i$ which garanties that we maintains the individual level of significance.

Lemma 4 *The distribution of Eq. 13 is stochastically decreasing in* $|\mu_i|$ *for all* $i \neq j$.

Proof: The proof follows from the stochastic ordering lemma and lemmas 1 and 2. **Q.E.D.**
The lemma tells us that we can construct the confidence intervals from the distribution of Eq. 13 by setting $\mu_i = 0$ for all i.

Theorem 2 *The confidence intervals for* μ_i *below keeps at least the confidence level* $100(1-\alpha)\%$*:*

$$x_i \pm z_\alpha(r,n)\hat{\sigma}_i$$

where α *is chosen such that* $\mathbb{P}\left\{X_i/\hat{\sigma}_i > z_\alpha(r,n)\right\} = \alpha$.

In table 2 the 0.05, 0.01 and 0.001 percentiles of the confidence interval-distribution is shown. The table is based on 10,000 simulations each. The confidence intervals will by necessity be conservative if large true effects are present. Assume that we have a 2^4 full factorial design and use the 8 smallest absolute values for estimating the variance and construct the confidence intervals. We further assume that there is one very large effect. This effect is so large that it in no way influences the variance estimation. Then the confidence intervals should rather be based on a table for $n = 14$ and $r = 8$ and hence the confidence intervals will be conservative.

5 Power simulations

The numerical study will limit itself to comparing with Voss (1999). In Voss (1999) one can find comparisons between GMR and other methods. If one

n	k	5%	1%	0.1%	n	k	5%	1%	0.1%
3	1	2.880	15.007	142.234	27	13	1.858	2.530	3.500
4	2	1.551	3.616	10.399	28	14	1.847	2.506	3.462
5	2	2.123	5.082	15.345	29	14	1.857	2.518	3.501
6	3	1.736	3.256	7.430	30	15	1.854	2.493	3.392
7	3	1.979	3.716	8.339	31	15	1.858	2.506	3.450
8	4	1.779	2.970	6.190	32	16	1.856	2.487	3.359
9	4	1.905	3.176	6.516	33	16	1.867	2.498	3.366
10	5	1.783	2.769	4.889	34	17	1.857	2.487	3.340
11	5	1.878	2.954	5.306	35	17	1.867	2.486	3.351
12	6	1.811	2.707	4.455	36	18	1.865	2.469	3.281
13	6	1.862	2.802	4.517	37	18	1.873	2.489	3.304
14	7	1.812	2.637	4.164	38	19	1.867	2.482	3.309
15	7	1.858	2.742	4.329	39	19	1.877	2.487	3.314
16	8	1.823	2.608	3.925	40	20	1.870	2.466	3.231
17	8	1.853	2.670	4.121	41	20	1.874	2.472	3.271
18	9	1.829	2.572	3.853	42	21	1.874	2.460	3.244
19	9	1.852	2.632	3.900	43	21	1.878	2.469	3.229
20	10	1.840	2.570	3.730	44	22	1.875	2.470	3.221
21	10	1.856	2.586	3.761	45	22	1.883	2.484	3.235
22	11	1.834	2.532	3.619	46	23	1.876	2.463	3.219
23	11	1.856	2.563	3.643	47	23	1.883	2.467	3.255
24	12	1.841	2.537	3.505	48	24	1.879	2.465	3.216
25	12	1.854	2.539	3.476	49	24	1.886	2.471	3.215
26	13	1.844	2.518	3.505	50	25	1.881	2.460	3.196

Table 2: Confidence intervals percentiles given n and k.

is interested in the performance of our method compared with others than GMR Voss (1999) is a good place to start.The GMR method is performed as follows: Decide on the number of estimates believed to be merely random noise. Let us say that there should be at least k such estimates X_i. We form the Quasi Mean Square Error (QMSE) according to

$$\text{QMSE} = \frac{1}{k}\sum_{i=1}^{k} X_{(i)}^2.$$

We now base the step down test on

$$T_j = \frac{X_{(j)}^2}{\text{QMSE}}.$$

The largest of these T_j:s namely T_1 is compared with a suitable critical value $c(1, n, k, \alpha)$. If $T_1 < c(1, n, k, \alpha)$ we make no rejections and stop. If $T_1 > c(1, n, k, \alpha)$ we reject the hypothesis that the largest effect is a null effect and continue to test if $T_2 > c(2, n, k, \alpha)$ etc.

The simulations are made by choosing a specific set of non-zero β_i and running repeated tests on these. Since we have a step down procedure the power for each individual hypothesis will be decreasing since for instance the second hypothesis can only be rejected if the first is rejected. We will count the number of correct rejections for the true effects and the number of cases where *a*ny false rejections are made. In all of the simulations we use $n = 15$ and $k = 8$ i.e. a total of 15 estimated effects from which the eight smallest are used for the variance estimation.

5.1 Four equal effects

We now use four equal effects all the same size as the noise level of experiment. If the standard deviation of the measurement error is σ then $\beta_i = \sigma$ for four randomly selected i:s. In this setting both methods have very little power. This is however natural since the true effect is small compared with the noise level. The estimated Type I error probability is also given. This will be smaller then the designated $\alpha = 0.05$ since our variance estimate is too big when true effects are present. This conservativeness is however smaller than that of the GMR testing procedure. This is probably due to the more exact variance estimation method. The method presented in this paper will be abbreviated CVMR (Censored Variance estimation Modulus Ratio). In

Method	H_{01}	H_{02}	H_{03}	H_{04}	$\mathbb{P}\{\text{Type I error}\}$
CVMR	0.0874	0.0378	0.0185	0.0099	0.01275
GMR	0.0932	0.0284	0.00865	0.00245	0.00735

Table 3: Estimated power and type I error probability

table 3 the estimated power for each of the possible true hypothesis is given together with the estimated probability of a type I error. The table is based on 20,000 simulation where four β_i were set to equal σ.

In the next simulation four β_i:s were set to 2σ. The result of this simulations is shown in table 4.

Method	H_{01}	H_{02}	H_{03}	H_{04}	$\mathbb{P}\{\text{Type I error}\}$
CVMR	0.411	0.312	0.260	0.188	0.03775
GMR	0.419	0.256	0.151	0.0723	0.0042

Table 4: Estimated power and type I error probability

5.2 Four unequal effects

For X_i we now put $\mu_i = i\sigma$ and do as in the previous section. The result of 40,000 simulations is shown in table 5. The table clearly shows that the

Method	H_{01}	H_{02}	H_{03}	H_{04}	$\mathbb{P}\{\text{Type I error}\}$
CVMR	0.933	0.764	0.410	0.0875	0.0314
GMR	0.931	0.661	0.237	0.0239	0.0051

Table 5: Estimated power and type I error probability

power is greater in the CVMR method than in the GMR method. This is probably due to the fact that the variance estimation is less conservative than the one used in GMR.

5.3 Conclusions concerning power

The CVMR method shows to have greater power when it comes to rejecting more than the first hypothesis. The variance estimation method used in CVMR uses information from the tail of the distribution in terms of the failure rate function. This stabilises the estimation and makes it more robust against the random censoring that the presence of true effects gives raise to. To gain more robustness against random censoring we need to use more advanced estimation methods that are adaptive. The use of adaptive methods for testing is an open problem when trying to maintain the correct multiple level of significance.

References

T. Adolfsson. *Using regression function information in determining the same.* PhD thesis, Chalmers University of Technology and Göteborg University, 2002.

K. N. Berk and R. R. Picard. Significans tests for saturated orthogonal arrays. *Journal of Quality Technology*, 1991.

G. E. P. Box, W. Hunter, and S. Hunter. *Statistics for experimenters.* Wiley, 1978.

C. Daniel. Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics*, 1959.

C. Daniel. *Applications of Statistics to Industrial Exprimentation.* Wiley, New York, 1976.

P. Haaland and M. O'Connel. Inference for effect-saturated fractional factorials. *Technometrics*, 1995.

E. L. Lehmann. *Testing Statistical Hypotheses.* Springer-Verlag, 1986.

D. Voss. Analysis of orthogonal saturated designs. *J. Stat. Plan. and Inf.*, 1999.

D.A. Zahn. An empirical study of the half-normal plot. *Technometrics*, 1975.