ON ADAPTIVE TESTING IN ORTHOGONAL
SATURATED DESIGNS

Daniel T. Voss and Weizhen Wang

Wright State University

Abstract: Adaptive, size-α step-down tests are provided for the analysis of orthogonal saturated designs. The tests work effectively under effect sparsity, and include as special cases the individual nonadaptive tests of Berk and Picard (1991) and the simultaneous nonadaptive tests of Voss (1988). The approach is similar to that used by Wang and Voss (2003) to construct adaptive confidence intervals, but testing is simpler because one can use the same denominator for all statistics. Step-down tests also have a clear power advantage over simultaneous confidence intervals and analogous single-step tests, as is demonstrated theoretically and assessed via simulation.

Key words and phrases: Closed test, effect sparsity, factorial design, step-down test, stochastic ordering.

1. Introduction

Consider the analysis of an orthogonal saturated factorial design involving k factorial effects $\beta_1, \ldots, \beta_k$ associated with multiple factors each at two levels. Assume a standard linear model for analysis of independent observations: $Y_i \sim N(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}, \sigma^2)$, $i = 1, \ldots, n$. The design is *orthogonal* if the least squares estimators $\hat{\beta}_i$ ($1 \leq i \leq k$) are uncorrelated, and the design is *saturated* if there are just enough observations to estimate the model parameters β_i (i.e., $n = k + 1$), leaving no error degrees of freedom to independently estimate the error variance σ^2. Of special interest is an orthogonal fraction of a 2^m factorial design that has just enough observations to estimate all main effects, since such designs have important applications in industry.

The analysis of orthogonal saturated designs was considered initially by Birnbaum (1959) and Daniel (1959), and subsequently by Zahn (1969, 1975a, 1975b). Then Box and Meyer (1986) presented a Bayesian method for the analysis of such data and Lenth (1989) proposed a "quick and easy (frequentist) analysis of unreplicated factorials" which was adaptive to the presence of some non-negligible effects. These later papers were apparently the catalyst for a flurry of research

on the analysis of orthogonal saturated designs, as many methods have been proposed since the late 1980's. See Hamada and Balakrishnan (1998) for a review and empirical comparison of such methods.

However, most of the proposed methods of analysis of saturated designs are only justified empirically. Relatively few of the methods have been shown to provide strong control of error rates — namely, control under all parameter configurations − espoused by Hochberg and Tamhane (1987, Chap.1) to be a fundamentally desirable property. For orthogonal saturated designs, nonadaptive methods of analysis known to provide strong control of error rates include step-down tests of Voss (1988), individual tests of Berk and Picard (1991), individual confidence intervals of Voss (1999), and simultaneous confidence intervals of Voss and Wang (1999). For nonorthogonal saturated designs, Kinateder, Voss and Wang (2000a) provided nonadaptive individual confidence intervals strongly controlling error rates, building upon a variance estimation approach of Kunert (1997). Kinateder, Voss and Wang (2000b) provided a review of methods of analysis of saturated designs known to strongly control error rates.

A related open problem of considerable interest is to show that step-down tests can be applied iteratively. For example Daniel (1959), in his consideration of subjective analysis based on half-normal plots, advocated redrawing the plot for the remaining estimators each time the largest estimator remaining under consideration was judged to be significantly large. Likewise, Voss (1988) observed that his step-down test would be more powerful if one used sharper critical values based on such an iterative approach. Venter and Steel (1998) advocated the use of this iterative approach, recognizing that it is justifiable if all effects are zero or infinite. Al-Shiha and Yang (1999) also proposed use of an iterative step-down test. Still, it remains open to show that iterative step-down tests strongly control error rates.

Yet another open problem of great interest is to show that adaptive methods, including Lenth's (1989) method, strongly control error rates. Lenth's method is adaptive in the sense that relatively large effect estimates $\hat{\beta}_i$ are set aside in the estimation of σ, and the number set aside depends on the estimates. More generally, a method is said to be *adaptive* if it uses one of several possible denominators and the choice is data dependent. Such adaptive estimation of variability is thought to be more efficient. Chen and Kunert (2004) proposed a procedure that is adaptive in this sense. Still, to show that such adaptive methods strongly control error rates remains open.

Ye, Hamada and Wu (2001) proposed the iterative, step-down application of Lenth's method, incorporating both of the aforementioned open problems into one methodology.

In this paper, adaptive step-down tests are proposed and shown to be of simultaneous size α. The tests are analogous to adaptive confidence intervals introduced recently by Wang and Voss (2001, 2003). However, the tests have two distinct advantages over the confidence intervals. First, while the confidence interval for each effect requires a different statistic denominator, such is not the case for the test statistics, simplifying test implementation. More importantly, step-down tests have an obvious power advantage, as will be shown in Section 3. Like the confidence intervals, the tests introduced here use critical values computed under the null parameter configuration (i.e., $\beta_i = 0, \forall\, i \geq 1$), and work effectively under *effect sparsity* — namely, when few of the effects β_i ($i \geq 1$) are non-negligible. The methodology and results are presented in Section 2, followed by consideration of power in Section 3.

2. Adaptive Tests

In this section, methods and results are presented for adaptively testing the hypotheses $H_{0i} : \beta_i = 0$ for $i = 1, \ldots, k$ either individually or simultaneously. The tests generalize the individual nonadaptive tests of Berk and Picard (1991) and the simultaneous nonadaptive tests of Voss (1988).

Assume the effect estimators $\hat{\beta}_i$ for $i \geq 1$ are independent with $\hat{\beta}_i \sim N(\beta_i, a_i^2 \sigma^2)$ for known constants a_i, and assume each $a_i = 1$ without loss of generality. These assumptions are standard for the analysis of orthogonal factorial designs. Let $K = \{1, \ldots, k\}$ denote the set of indices corresponding to individual hypotheses $H_{0i} : \beta_i = 0$. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ and $\boldsymbol{\beta}_{null} = (0, \ldots, 0)$.

Stochastic ordering plays a fundamental role in establishing test size. The distribution of $X \sim F_\theta(x)$ is said to be *stochastically nondecreasing* in θ if $P_\theta(X > x) = 1 - F_\theta(x)$ is nondecreasing in θ for all x. The following stochastic ordering lemma of Alam and Rizvi (1966) and Mahamunulu (1967) is simple but quite useful.

Lemma 1. *Let $X_1, \ldots, X_k$ be independent random variables, where $X_i \sim F_{\theta_i}(x_i)$ is stochastically nondecreasing in θ_i. If the statistic $t = t(x_1, \ldots, x_k)$ is a nonincreasing function of x_i for each i when all x_j ($j \neq i$) are held fixed, then the distribution of $T = t(X_1, \ldots, X_k)$ is stochastically nonincreasing in θ_i.*

The closed testing procedure of Marcus, Peritz, and Gabriel (1976) is the standard approach for establishing simultaneous step-down tests of size α. Their procedure requires specification of a size-α test of the hypothesis $H_{0I} : \beta_i = 0 \,\forall\, i \in I$ for each nonempty $I \subset K$. To test H_{0I}, consider the test statistic

$$T_I = \max_{i \in I} T_i, \quad \text{for } T_i = \frac{\hat{\beta}_i^2}{\min_{j \in J}\{c_j\, qmse_j\}}, \tag{1}$$

where the *quasi mean squared error* $qmse_j = (1/j)\sum_{h=1}^{j} \hat{\beta}_{(h)}^2$ is the average of the j smallest of the k order statistics $\hat{\beta}_{(1)}^2 \leq \cdots \leq \hat{\beta}_{(k)}^2$ of the squared estimators, $J \subset K$ is a predetermined nonempty subset of K, and the c_j are specified positive constants. For example, for testing 15 effects, consider using $J = \{8, 12\}$ with the constants c_j ($j = 8, 12$) chosen so $E_{\boldsymbol{\beta}_{null}}[c_j QMSE_j] = \sigma^2$. Then the denominator pools together either the eight or twelve smallest squared estimates, favoring the use of eight, for example, if $\hat{\beta}_{(12)}^2$ is large.

Let $t_{|I|,\alpha}$ denote the upper-α quantile of the distribution of the test statistic T_I at $\boldsymbol{\beta}_{null}$; i.e.,

$$P_{\boldsymbol{\beta}_{null}}(T_I > t_{|I|,\alpha}) = \alpha. \tag{2}$$

Here $|I|$ denotes the cardinality of the index set I. The distribution of T_I is nonstandard, but the critical value $t_{|I|,\alpha}$ is easily approximated via simulation. For a subset $I' \subset I$, it is obvious that $T_{I'} \leq T_I$. Therefore, $t_{|I'|,\alpha} \leq t_{|I|,\alpha}$.

Lemma 2. *For a nonempty subset I of K, a size-α test of H_{0I} is to reject H_{0I} if $T_I > t_{|I|,\alpha}$.*

Proof. Let $\boldsymbol{\beta}_{0I} = \{(\beta_1, \ldots, \beta_k) : \beta_i = 0 \text{ for } i \in I\}$. It suffices to show that $P_{\boldsymbol{\beta}_{0I}}(T_I > t_{|I|,\alpha}) \leq \alpha$ over $\boldsymbol{\beta}_{0I}$. However, T_I is a nonincreasing function of $\hat{\beta}_h^2$ for each $h \notin I$ and $\hat{\beta}_h^2$ is stochastically nondecreasing in β_h^2, so T_I is stochastically nonincreasing in β_h^2 by Lemma 1.

Corollary 1. *For an integer i between 1 and k, a size-α test of $H_{0i} : \beta_i = 0$ is to reject H_{0i} if $T_i > t_{1,\alpha}$.*

A more important consequence of Lemma 2 is the following. The closed test for simultaneously testing the hypotheses $H_{0i} : \beta_i = 0$ ($i \in K$) controls the probability of making *any* false assertions to be at most α. This closed test, which rejects H_{0i} if H_{0I} is rejected for every $I \subset K$ containing i, controls the simultaneous error rate (Marcus, Peritz, and Gabriel, (1976)) but is rather cumbersome to conduct. However, for closed tests of this type, the following step-down shortcut to the closed test also controls the error rate, (Hochberg and Tamhane, (1987, p.55)).

Step-down Tests: Let $[1], \ldots, [k]$ be random indices such that $T_{[1]} < \cdots < T_{[k]}$.

  Step 1: If $T_{[k]} > t_{k,\alpha}$, then infer $\beta_{[k]} \neq 0$ and continue; else stop.
  Step 2: If $T_{[k-1]} > t_{k-1,\alpha}$, then infer $\beta_{[k-1]} \neq 0$ and continue; else stop.

$$\vdots$$

Iterate in this step-down fashion, but stop the first time no nonzero effect is claimed. Under effect sparsity, the procedure typically stops within a few steps.

Theorem 1. *For the above step-down test of $H_{0i} : \beta_i = 0$ $(i = 1, \ldots, k)$, the probability of making any false inferences is at most α for any $\boldsymbol{\beta}$. That is,*

$$\max_{I \subset K} \sup_{\boldsymbol{\beta} \in H_{0I}} P_{\boldsymbol{\beta}}(reject \ H_{0i} \ for \ some \ i \in I) \leq \alpha.$$

Remarks. If $|J| = 1$ in equation (1), then the tests of Lemma 2, Corollary 1 and Theorem 1 are nonadaptive. In this case, the individual tests of Corollary 1 reduce to the nonadaptive tests of Berk and Picard (1991), and the simultaneous step-down test reduces to the nonadaptive, closed, step-down test of Voss (1988).

However, if $|J| > 1$ and $c_j > c_{j'}$ for $j < j'$ $(j, j' \in J)$, then these same tests are *fully adaptive* in the following sense. There is positive probability that the denominator $\min_{j \in J}\{c_j qmse_j\}$ of each test statistic T_i in equation (1) will be equal to $c_j qmse_j$ for each $j \in J$.

In particular, the tests of Lemma 2, Corollary 1 and Theorem 1 are fully adaptive if one chooses the constants c_j in equation (1) to make $E_{\boldsymbol{\beta}_{null}}[c_j \, QMSE_j] = \sigma^2$ for each $j \in J$. This approach to choosing the constants c_j was found to be efficient in simulations conducted by Wang and Voss (2003) in their consideration of adaptive confidence intervals. The denominators they considered were analogous except, in getting a confidence interval for β_i, the denominator is kept free of $\hat{\beta}_i$.

The form of the adaptive statistic T_i in equation (1) is also motivated by the following fact. The nonadaptive statistic $T_i = \hat{\beta}_i^2/qmse_\nu$, corresponding to $J = \{\nu\}$ with $c_\nu = 1$, would be the likelihood ratio statistic for testing $H_{0i} : \beta_i = 0$ if one knew that exactly ν of the effects β_j were zero without knowing specifically which ones, (Al-Shiha and Yang (1999)).

3. Advantage of Stepping Down

In the step-down tests of Section 2, the test statistics $T_{[k]}$, $T_{[k-1]}, \ldots$ are compared to the respective critical values in the decreasing sequence $\{t_{i,\alpha}\}_{i=k}^1$. This is in contrast to a single-step simultaneous size-α test that compares each test statistic T_i $(i = 1, \ldots, k)$ to the largest critical value, $t_{k,\alpha}$, and asserts $\beta_i \neq 0$ if $T_i > t_{k,\alpha}$. Hence the simultaneous step-down tests, justified by the closure method, have an obvious power advantage over the single-step test. The critical values for the step-wise tests are sharper for testing each effect except the one with the largest estimate, and the critical values are the same for testing that effect. Hence, any effects asserted nonzero by the single-step test are also asserted nonzero by the corresponding step-down test, though the converse need not be true. One would like to have step-wise confidence intervals or sets analogous to the step-down tests. Hsu and Berger (1999) have obtained step-wise confidence

intervals when it is desirable to give inferences in a specified order. Their justi-
fication is based on the so-called partitioning principle, involving a partitioning
of the parameter space. However, in the setting considered here and by Wang
and Voss (2003), there is no natural order in which to consider the parameters
β_i, so the approach of Hsu and Berger (1999) does not apply. Rather, in this
setting, simultaneous confidence intervals are analogous to single-step tests, so
are similarly at a power disadvantage compared to step-down tests. This is the
reason step-down tests are useful for analysis of orthogonal saturated designs.

   In this section, the nature and magnitude of this power advantage is inves-
tigated via simulation. Power was estimated via simulation for the case of 15
estimators, as one would have in the analysis of an orthogonal fractional fac-
torial design for estimating 15 effects in 16 runs. Included were six parameter
configurations, including from four to six nonzero effects of size varying from one
to five standard deviations of the data. Fewer nonzero effects yield less power
advantage, and there is no advantage for only one nonzero effect.

   For each of six parameter configurations $\boldsymbol{\beta}$, 10,000 parameter vector esti-
mates $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_{15})$ were generated assuming independent estimates $\hat{\beta}_i \sim$
$N(\beta_i, \sigma^2)$ for $i = 1, \ldots, 15$, taking $\sigma = 1$. Each vector $\hat{\boldsymbol{\beta}}$ of estimates was
analyzed using the step-down test procedure of Section 2, and with the corre-
sponding single-step test. In both cases, the same denominator of T_i was used—
namely, $\min_{j \in J}\{c_j qmse_j\}$, with $J = \{8, 12\}$, and with coefficients c_j chosen so
$E_{\boldsymbol{\beta}_{null}}[c_j QMSE_j] = \sigma^2$.

   The simulation results are summarized in Table 1. The first configuration
considered has the five nonzero effects $\beta_i = i$ for $i = 1, \ldots, 5$, with the remaining
effects zero, (i.e. $\beta_i = 0$ for $i = 6, \ldots, 15$). The effect size is in error standard
deviations σ, assuming each estimator $\hat{\beta}_i$ has variance $\sigma^2/4$. Test power is com-
puted separately for each effect size. One can see that both the step-down and
the single-step procedures asserted the effect of size five (i.e., here $\beta_5 = 5$) to
be nonzero with power 0.978. This is not surprising since both procedures use
the same critical value in the first step, though the step-down procedure can
have slightly enhanced power even for the largest effect since the largest effect
need not always correspond to the largest estimator. Also not surprisingly, the
step-down procedure gains power in testing the smaller effects, with for example
a 14% gain in power for the effect of size two.

   Perhaps the most striking results are obtained for the fourth configuration
listed, for which $\beta_1 = \beta_2 = 3$, $\beta_3 = \beta_4 = 4$ and $\beta_5 = \beta_6 = 5$. In this case, for the
two effects of size three, the gain in power from stepping down is a substantial
20% compared to using the single-step procedure.

   As one can see from the table, power gains provided by the step-down pro-
cedure are better when effect sizes vary than when all nonzero effects are of the
same size, such as in the last case with six nonzero effects each of size five. It is
the power to detect the smaller effects which is most enhanced.

Table 1. Power comparison for step-down and single-step methods.

Configuration	Method	Effect Size				
$\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 3$,		1	2	3	4	5
$\beta_4 = 4$, $\beta_5 = 5$,	Step-down	0.015	0.175	0.547	0.866	0.978
$\beta_i = 0$ otherwise	Single-step	0.012	0.153	0.519	0.856	0.978
	Percent Gain	25%	14%	5%	1%	0%
$\beta_1 = 2$, $\beta_2 = 3$,		2	3	4	5	
$\beta_3 = 4$, $\beta_4 = 5$,	Step-down	0.248	0.662	0.923	0.991	
$\beta_i = 0$ otherwise	Single-step	0.218	0.632	0.918	0.991	
	Percent Gain	14%	5%	0.5%	0%	
$\beta_1 = \beta_2 = 2$,		2	4			
$\beta_3 = \beta_4 = 4$,	Step-down	0.240	0.923			
$\beta_i = 0$ otherwise	Single-step	0.218	0.919			
	Percent Gain	10%	0.4%			
$\beta_1 = \beta_2 = 3$,		3	4	5		
$\beta_3 = \beta_4 = 4$,	Step-down	0.375	0.711	0.915		
$\beta_5 = \beta_6 = 5$,	Single-step	0.313	0.672	0.909		
$\beta_i = 0$ otherwise	Percent Gain	20%	6%	0.7%		
$\beta_1 = \cdots = \beta_6 = 3$,		3				
$\beta_i = 0$ otherwise	Step-down	0.341				
	Single-step	0.313				
	Percent Gain	9%				
$\beta_1 = \cdots = \beta_6 = 5$,		5				
$\beta_i = 0$ otherwise	Step-down	0.933				
	Single-step	0.908				
	Percent Gain	3%				

Acknowledgements

References

Alam, K. and Rizvi, M. H. (1966). Selection from multivariate normal populations. *Ann. Inst. Statist. Math.* **18**, 307–318.

Al-Shiha, A. A. and Yang, S. S. (1999). A multistage procedure for analyzing unreplicated factorial experiments. *Biometrical J.* **41**, 659-670.

Berk, K. N. and Picard, R. R. (1991). Significance tests for saturated orthogonal arrays. *J. Qual. Tech.* **23**, 79-89.

Birnbaum, A. (1959). On the analysis of factorial experiments without replication. *Technometrics* **1**, 343-357.

Box, G. E. P. and Meyer, R. D. (1986). An analysis for unreplicated fractional factorials. *Technometrics* **28**, 11-18.

Chen, Y. and Kunert, J. (2004). A new quantitative method for analysing unreplicated factorial designs. *Biometrical J.* **46**, 125-140.

Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics* **1**, 311-341.

Hamada, M. and Balakrishnan, N. (1998). Analyzing unreplicated factorial experiments: A review with some new proposals. *Statist. Sinica* **8**, 1-41.

Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*. Wiley, New York.

Hsu, J. C. and Berger, R. L. (1999). Stepwise confidence intervals without multiplicity adjustment for dose-response and toxicity studies, *J. Amer. Statist. Assoc.* **94**, 468-482.

Kinateder, K. K. J., Voss, D. T. and Wang, W. (2000a). Exact confidence intervals in the analysis of nonorthogonal saturated factorial designs, *Amer. J. Math. Management Sci.* **20**, 71-84.

Kinateder, K. K. J., Voss, D. T. and Wang, W. (2000b). Analysis of saturated and super-saturated factorial designs: a review. In *Proceedings of the Indian International Statistical Association* 1998 *International Conference* (Edited by N. Balakrishnan), 325-347. Gordon and Breach, New Jersey.

Kunert, J. (1997). On the use of the factor-sparsity assumption to get an estimate of the variance in saturated designs. *Technometrics* **39**, 81-90.

Lenth, R. V. (1989). Quick and easy analysis of unreplicated factorials. *Technometrics* **31**, 469-473.

Mahamunulu, D. M. (1967). Some fixed-sample ranking and selection problems. *Ann. Math. Statist.* **38**, 1079-1091.

Marcus, R., Peritz, E. and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63**, 655-660.

Venter, J. H. and Steel, S. J. (1998). Identifying active contrasts by stepwise testing. *Technometrics* **40**, 304-313.

Voss, D. T. (1988). Generalized modulus-ratio tests for analysis of factorial designs with zero degrees of freedom for error. *Comm. Statist. Theory Methods* **17**, 3345-3359.

Voss, D. T. (1999). Analysis of orthogonal saturated designs. *J. Statist. Plann. Inference* **78**, 111-130.

Voss, D. T. and Wang, W. (1999). Simultaneous confidence intervals in the analysis of orthogonal saturated designs. *J. Statist. Plann. Inference* **81**, 383-392.

Wang, W. and Voss, D. T. (2001). Control of error rates in adaptive analysis of orthogonal saturated designs. *Ann. Statist.* **29**, 1058-1065.

Wang, W. and Voss, D. T. (2003). On adaptive estimation in orthogonal saturated designs. *Statist. Sinica* **13**, 727-737.

Ye, K. Q., Hamada, M. and Wu, C. F. J. (2001). A step-down Lenth method for analyzing unreplicated factorial designs. *J. Qual. Technol.* **33**, 140-152.

Zahn, D. A. (1969). An empirical study of the half-normal plot. Ph.D. thesis, Harvard Univ.

Zahn, D. A. (1975a). Modifications of and revised critical values for the half-normal plots. *Technometrics* **17**, 189-200.

Zahn, D. A. (1975b). An empirical study of the half-normal plot. *Technometrics* **17**, 201-211.

Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435, U.S.A.

E-mail: dan.voss@wright.edu

Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435, U.S.A.

E-mail: wwang@math.wright.edu