ESTABLISHING STRONG CONTROL OF FAMILYWISE ERROR RATE FOR ITERATIVE STEP-DOWN TESTS

By DANIEL T. VOSS and WEIZHEN WANG
Wright State University

E04-154 rev2, April 24, 2006

Abstract

A long-standing problem concerning the iterative analysis of orthogonal saturated designs has been resolved. Consider an unreplicated factorial design yielding independent, normally distributed estimators of k parameters but no independent variance estimator. Many authors have proposed iterative step-down tests for the analysis of such designs. In essence, the first such methods were proposed by Birnbaum [3] and Daniel [4]. It is well known that iterative methods are more powerful than corresponding closed step-down tests (see Voss [12]), and so the iterative tests are more popular. This popularity has grown despite lack of a proof that the iterative methods strongly control the familywise error rate, while some corresponding closed step-down tests with modestly less power have been known to provide such control. Venter and Steele [11] claimed that certain iterative step-down tests strongly control error rates, but they failed to provide a proof. Recently Holm, Mark and Adolfsson [5] provided the first iterative step-down test for analysis of orthogonal saturated designs shown to strongly control the familywise error rate. Using the same technical

[1] *AMS 2000 subject classifications.* Primary 62K15, 62F07, 62F03; secondary 62F35, 62L10.

[2] *Key words and phrases.* Adaptive test, effect sparsity, factorial experiment, multiple test, orthogonal saturated design, stochastic order, nearly saturated, unreplicated factorial.

[3] Acknowledgment: This research was supported by NSF through grant number DMS0308861.

[4] *Abbreviated Title.* Iterative Step-Down Tests.

approach, but without the need for explicit consideration of coverage bounds, we establish strong control of the familywise error rate for a large class of iterative step-down tests, including the iterative tests of Zahn [19, 20] and Ventor and Steel [11], iterative variations on the tests of Daniel [4], Birnbaum [3], Voss [12], Voss and Wang [14], Lenth [8] and Ye, Hamada and Wu [18], and a generalization of the Holm, Mark and Adolfsson [5] test for orthogonal saturated designs. Also included are the iterative step-down tests of Langsrud and Naes [6] for nearly saturated designs. Since our approach does not explicitly involve coverage bounds, implementation is relatively simple. Also, the normality assumption can be relaxed.

1 Introduction

The analysis of data collected using an orthogonal saturated design in an important problem in statistics, since such designs include commonly used orthogonal fractional factorial designs. Building upon a specific step-down test and result of Holm, Mark and Adolfsson [5], we establish strong control of the familywise error rate for a large class of iterative step-down tests. This resolves an important and long-standing problem.

We now pose the problem as it typically arises—namely, for normally distributed estimators—though strong control of error rates is established in Section 2 for a broader class of distributions.

The standard problem is as follows. Let $\hat{\theta}_i \sim N(\theta_i,\, a_i^2\sigma^2)$, $i = 1, \ldots, k$, be k independent normal estimators, with constants a_i known but σ^2 unknown. The methods to be considered work best under effect sparsity—namely, when most of the effects θ_i are zero—but in practice one knows neither how many effects are zero nor which ones. Complicating matters, saturated designs provide no error degrees of freedom, so assume there is no independent variance estimator. The objective is to simultaneously test the hypotheses $H_{0i} : \theta_i = 0$ for $i = 1, \ldots, k$, and we require a testing procedure that strongly controls the simultaneous or familywise error rate—namely, that controls the probability of making any false assertions to be at most α under all parameter configurations $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Henceforth assume without loss of generality that each $a_i = 1$; otherwise one could equivalently consider estimators $\hat{\theta}_i/a_i \sim N(\theta_i/a_i,\, \sigma^2)$.

Daniel [4] and Birnbaum [3] were perhaps the first to consider this problem. Daniel proposed the iterative use of half-normal plots. Given the half-normal plot of the k effect estimates, if the effect with largest estimate is deemed to be significantly nonzero, Daniel advocated setting aside that estimate to evaluate the rest. Thus, one would redraw the half-normal plot for the remaining $k-1$ estimates, and use this plot to assess significance of the largest of the remaining $k-1$ estimates as if there had only been $k-1$ estimates. This process is iterated until the largest remaining estimate is not considered significantly nonzero. Daniel [4] and Birnbaum [3] also each discussed tests for one nonzero effect, and these tests could be applied iteratively.

In efforts to formalize the subjective graphical method of Daniel [4], various step-down tests have been proposed, including tests by Zahn [19, 20], Voss [12], Ventor and Steel [11], Langsrud and Naes [6], Al-Shiha and Yang [2], Ye, Hamada and Wu [18], and Voss and Wang [14].

Voss [12] provided the first step-down test known to strongly control the familywise error rate in this setting, and his method was generalized by Voss and Wang [14] to include adaptive tests. These tests strongly control error rates because they are step-down shortcuts to closed tests, closed tests having been introduced by Marcus, Peritz, and Gabriel [9]. Voss observed that his test would have more power if applied iteratively, but he was unable to establish strong control of the familywise error rate for the iterative version of his procedure. Iterative testing has also been advocated by Zahn [19, 20], Ventor and Steel [11], and Al-Shiha and Yang [2]. Ventor and Steel [11] proposed the use of iterative step-down tests for enhanced power over certain closed tests. They claimed their iterative step-down tests provided strong control of the simultaneous error rate, but they failed to provide a proof. They did observe that their claim was true if the effects were all zero or infinite, but it has remained an open problem to establish this result for all parameter configurations. Ye, Hamada and Wu [18] proposed applying the adaptive test of Lenth [8] iteratively and stepping down, but it remains open to show that the tests of Lenth [8] or Ye, Hamada and Wu [18] strongly control the familywise error rate because of a lack of monotonicity in the way their tests are adaptive.

For the various multiple testing procedures proposed for this setting, critical values are typically computed under the *(complete) null distribution*—namely, when all k effects are zero—and the usual presumption is that this is sufficient to control the probability of making any false assertions under all parameter configurations. This has remained an open problem however, except in the case of closed step-down tests. When the step-down shortcut for a closed test is valid, error rate is still strongly controlled, due in part to the fact that each critical value in the closed step-down test is computed using k estimators under the null distribution. In contrast, when the maximum estimate is tested iteratively, each time an assertion is made the next critical value is computed using one less estimator. Consequently, critical values are sharper so power is increased, but the test is not closed so error rate control needs to be established in another fashion.

Holm, Mark and Adolfsson [5] provided a new step-down test for analysis of orthogonal saturated designs that is remarkable for a number of reasons. The test dovetails very nicely with the half-normal plot, which is commonly used for the subjective analysis of the estimates in this setting. Their test is an adaptive procedure that provides strong familywise control of error rates, and the only other inference procedures known to be adaptive and provide strong familywise control of error rates are the confidence intervals of Wang and Voss [15, 16] and corresponding tests of Voss and Wang [14]. The procedure is based on coverage bounds for order statistics of the estimates, an approach to analysis of such data that has perhaps not been considered since Birnbaum [3]. Most importantly for the purpose of this paper, Holm, Mark and Adolfsson [5] provide a new line of proof that their step-down test controls the familywise error rate over all parameter configurations. It is this line of proof, sketched in their appendix, that we exploit in the current paper to resolve a long-standing problem. Specifically, using the same technical approach, we establish strong control of the familywise error rate for a much larger class of iterative step-down tests for analysis of orthogonal saturated designs, (see Section 2, Theorem 1). Furthermore, the result is obtained without explicit consideration of coverage bounds, greatly simplifying implementation. Then in Section 3 we apply Theorem 1 to establish strong control of the familywise error rate for several iterative step-down tests and iterative versions of other tests proposed in the literature. Methods covered for the analysis of orthogonal saturated designs include the iterative step-down tests of Zahn [19, 20] and Venter and Steel [11], iterative versions of tests proposed by Daniel [4], Birnbaum [3], Voss [12], Voss and Wang [14], iterative variations on the adaptive methods proposed by Lenth [8] and Ye, Hamada and Wu [18], as well as a generalization of the method of Holm, Mark and Adolfsson [5]. Also covered is the iterative step-down test of Langsrud and Naes [6] for the analysis of nearly-saturated orthogonal designs.

2 Step-down tests

In this section we establish strong control of the familywise error rate for a large class of iterative step-down tests, including that of Holm, Mark and Adolfsson [5], but relaxing the usual normality assumption. First we introduce some notation.

Let $X_i \sim (1/\sigma)f(x_i/\sigma;\theta_i/\sigma)$, $i = 1,\ldots,k$, be independently distributed random variables, for $f(t;\theta_i)$ a pdf with positive support. For example, in the standard setting, X_i is typically the absolute value or square of the ith of k normally distributed estimators $\hat{\theta}_i$, each scaled to have common variance σ^2. Let $[1],[2],\ldots,[k]$ be the random indices such that $X_{[1]} < X_{[2]} < \cdots < X_{[k]}$.

Consider as test statistics

$$T_i = \frac{X_{[i]}}{g_i(X_{[1]},\ldots,X_{[i]})}, \quad i = 2,\ldots,k, \tag{1}$$

where the $g_i(t_1,\ldots,t_i)$, $i = 2,\ldots,k$, are positive functions with the following properties.

(i) (scale-equivariant) $g_i(at_1,...,at_i) = ag_i(t_1,...,t_i)$ for any $a > 0$;

(ii) (exchangable) $g_i(t_1,...,t_i) = g_i(t_{j_1},...,t_{j_i})$ for any permutation $(j_1,...,j_i)$ of $(1,...,i)$;

(iii) (non-decreasing) $s_i \leq t_i$ for $i = 1,\ldots,k$ implies $g_i(s_1,...,s_i) \leq g_i(t_1,...,t_i)$.

Obtain the critical values c_i, $i = 2,\ldots,k$, as follows. Let $Y_1,\ldots,Y_i$ be independent, each with pdf $f(y;0)$. For the corresponding order statistics $Y_{(1)} < \cdots < Y_{(i)}$, let c_i be the upper-α quantile of the distribution of

$$C_i = Y_{(i)}/g_i(Y_{(1)},\ldots,Y_{(i)}).$$

Note that C_i is a function of i order statistics, not k. Consider the following test procedure.

Iterative step-down test.

Step 1.	If $T_k \le c_k$ then make no assertions and stop; else assert $\theta_{[k]} \ne 0$ and continue.
Step 2.	If $T_{k-1} \le c_{k-1}$ then stop; else assert $\theta_{[k-1]} \ne 0$ and continue.
$\vdots$	$\vdots$
Step $k-1$.	If $T_2 \le c_2$ then stop; else assert $\theta_{[2]} \ne 0$ then stop.

So, if the procedure stops at step $j < k-1$, then exactly $j-1$ effects are asserted to be nonzero, and no other assertions are made. If it stops at step $k-1$, either $k-2$ or $k-1$ assertions are made, depending on whether one asserts $\theta_{[2]} \ne 0$. Clearly, $\theta_{[1]} \ne 0$ is never asserted.

Additional notation is needed for the following theorem. For $X_\theta \sim f(x;\theta)$ and fixed $c > 0$, let $X_{\theta c}$ have the conditional distribution of X_θ/c given $X_\theta \le c$. Also, let $U \preceq V$ denote that U is stochastically less than or equal to V, or equivalently, that $F_U(x) \ge F_V(x)$ for all x, where F_U is the cdf of U.

Theorem 1 *Let $X_i \sim (1/\sigma)f(x_i/\sigma;\theta_i/\sigma)$ be independent with positive support, for $i = 1,\ldots,k$. Furthermore, for $X_\theta \sim f(x;\theta)$, suppose:*

(i) $X_{0c} \preceq X_{\theta c}$ for any $\theta \ne 0$ and $c > 0$, and

(ii) $X_{0d} \preceq X_{0c}$ for any $d > c > 0$.

The above iterative step-down test strongly controls the probability of making any false assertions to be at most α under all parameter configurations $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_k)$ and for all σ.

Proof of the theorem uses the following lemma.

Lemma 1 *Suppose $h(t_1,...,t_k)$ is an exchangable function and $U_i \preceq V_i$ $(i = 1,...,k)$ are independent. If $h(t_1,...,t_k)$ is nonincreasing, then $h(V_1,\ldots,V_k) \preceq h(U_1,\ldots,U_k)$.*

Lemma 1 was independently discovered by Alam and Rizvi [1] and Mahamunulu [10], rediscovered by Voss [13], and stated in essentially the above form by Wu and Wang [17].

Lemma 2 *(Lehmann [7], p. 112) If $U \preceq V$, then $E[h(U)] \le E[h(V)]$ for any nondecreasing function $h()$.*

Proof of Theorem. The test statistics T_i in equation (1) are scale-invariant, so without loss of generality assume $\sigma = 1$. Consider any parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, and corresponding independent random variables $X_i \sim f(x_i; \theta_i)$, $i = 1, \ldots, k$. Let $I = \{1, \ldots, k\}$ denote the index set for all k effects. Assume at least one θ_i is zero; otherwise the theorem follows trivially. Let $I_0 = \{i_j : \theta_{i_j} = 0, j = 1, \ldots, m\}$ be the nonempty set of indices of the null effects for given $\boldsymbol{\theta}$, so $m = |I_0|$ denotes the number of null effects, $m \geq 1$. Also for any nonnegative integer n, let $\boldsymbol{Y} = (Y_1, \ldots, Y_{m+n-1})$ be a vector of i.i.d. random variables, each Y_i with pdf $f(y; 0)$, with corresponding order statistics $Y_{(1)} < \cdots < Y_{(m+n-1)}$. Let $\boldsymbol{Y}|c$ denote the distribution of $\boldsymbol{Y}$ given $Y_i < c$ for $i = 1, \ldots, m+n-1$. Also, let $Y_{(m+n)}$ denote the maximum of $m+n$ i.i.d. random variables, each with pdf $f(y; 0)$.

Let $Z = \max_{i \in I_0} X_i$ denote the largest estimator of the m null effects, N the number of non-null effects with estimate less than Z, and $\omega = \{i \in I : X_i < Z\}$, so $\omega = \{[1], \ldots, [m+N-1]\}$. Note that Z, N and ω are random and unobservable. Clearly, $Z = X_{[m+N]}$. Let $p_{\boldsymbol{\theta}}(n, \omega)$ denote the joint probability mass function of (N, ω). Also, let $\phi(x) = I_{\{x>0\}}$ denote an indicator function. Then

$$
\begin{aligned}
&P_{\boldsymbol{\theta}}(\text{any false assertions}) = P_{\boldsymbol{\theta}}(\text{assert } \theta_{i_j} \neq 0 \text{ for some } j \leq m) \\
&\quad \leq \; P_{\boldsymbol{\theta}}\left[Z/g_{m+N}(X_{[1]}, \ldots, X_{[m+N-1]}, Z) > c_{m+N}\right] \\
&\quad = \; \sum_{n=0}^{k-m} \sum_{\omega} p_{\boldsymbol{\theta}}(n, \omega) \times P_{\boldsymbol{\theta}}\left[Z/g_{m+N}(X_{[1]}, \ldots, X_{[m+N-1]}, Z) > c_{m+n} \mid N, \omega\right] \\
&\quad = \; \sum_{n=0}^{k-m} \sum_{\omega} p_{\boldsymbol{\theta}}(n, \omega) \times E_Z\left\{E_{\boldsymbol{X}|Z,N,\omega;\boldsymbol{\theta}}\left[\phi()\right]\right\}, \\
&\qquad\quad \text{where } \phi() = \phi\left(g_{m+n}(X_{[1]}/Z, \ldots, X_{[m+n-1]}/Z, 1)^{-1} - c_{m+n}\right) \\
&\quad \leq \; \sum_{n=0}^{k-m} \sum_{\omega} p_{\boldsymbol{\theta}}(n, \omega) \times E_Z\left\{E_{\boldsymbol{Y}|Z,N,\omega}\left[\phi()\right]\right\}, \\
&\qquad\quad \text{where } \phi() = \phi\left(g_{m+n}(Y_1/Z, \ldots, Y_{m+n-1}/Z, 1)^{-1} - c_{m+n}\right) \\
&\quad = \; \sum_{n=0}^{k-m} p_{\boldsymbol{\theta}}(n) \times E_Z\left\{E_{\boldsymbol{Y}|Z,N}\left[\phi()\right]\right\}, \\
&\quad \leq \; \sum_{n=0}^{k-m} p_{\boldsymbol{\theta}}(n) \times E_{Y_{(m+n)}}\left\{E_{\boldsymbol{Y}|Y_{(m+n)},N}\left[\phi()\right]\right\}, \\
&\qquad\quad \text{where } \phi() = \phi\left(g_{m+n}(Y_1/Y_{(m+n)}, \ldots, Y_{m+n-1}/Y_{(m+n)}, 1)^{-1} - c_{m+n}\right)
\end{aligned}
$$

$$= \sum_{n=0}^{k-m} p_{\boldsymbol{\theta}}(n) \times E_{(Y_{(1)},\ldots,Y_{(m+n)})}\left[\phi()\right] ,$$

$$\text{where } \phi() = \phi\left(Y_{(m+n)}/g_{m+n}(Y_{(1)},\ldots,Y_{(m+n-1)},Y_{(m+n)}) - c_{m+n}\right)$$

$$= \sum_{n=0}^{k-m} p_{\boldsymbol{\theta}}(n) \times \alpha \;=\; \alpha$$

The first inequality follows because $Z/g_{m+N}(X_{[1]},\ldots,X_{[m+N-1]},Z) > c_{m+N}$ is necessary for any false assertion to occur.

The second inequality follows from Lemma 1 and condition (i) of the theorem because, for a given (Z,N,ω), $\phi()$ is a non-increasing function of X_i/Z for each $i \in \omega$, which contains $m+n-1$ elements, each X_i/Z (given $X_i < Z$) is stochastically smallest at $\theta_i = 0$ by (i), the X_i's are independent, and $g()$ is exchangable.

The third inequality is a consequence of Lemmas 1 and 2 and condition (ii) of the theorem, as follows. For a given (N,Z), and given $Y_i < Z$ for $i = 1,\ldots,m+n-1$, $\phi()$ is a non-increasing function of the Y_i/Z, which are independent, and which are stochastically non-increasing in Z by condition (ii), so $\phi()$ is stochastically non-decreasing in Z by Lemma 1. Thus, $E_{\mathbf{Y}|Z}\phi()$ is non-decreasing in Z. Since $Z \preceq Y_{(m+n)}$, the inequality follows from Lemma 2, □

Theorem 1 will now be shown to include the standard cases under normality.

Lemma 3 *(Holm, Mark and Adolfsson* [5], *Wu and Wang* [17]*. Let $X_\theta = |\theta + Z\sigma|$ (a folded normal random variable) for $Z \sim N(0,1)$ and $\sigma > 0$ or let $X_\theta = (\theta + Z\sigma)^2$. Then*

(i) $X_{0c} \preceq X_{\theta c}$ for any $\theta \neq 0$ and $c > 0$, and

(ii) $X_{0d} \preceq X_{0c}$ for any $d > c > 0$.

Corollary 1 *Theorem 1 holds for $X_i = |\hat{\theta}_i/a_i|$ or $X_i = (\hat{\theta}_i/a_i)^2$, if $\hat{\theta}_1,\ldots,\hat{\theta}_k$ are independent with $\hat{\theta}_i \sim N(\theta_i, a_i^2\sigma^2)$, for known nonzero constants a_i. (The critical value c_i corresponding to T_i is obtained using independent folded standard normal or χ_1^2 random variables $Y_1,\ldots,Y_i$.)*

In the above corollary, the estimators $\hat{\theta}_i/a_i \sim N(\theta_i/a_i, \sigma^2)$ are necessarily scaled to have common variance, but testing $\theta_i/a_i = 0$ is equivalent to testing $\theta_i = 0$.

The proof of Theorem 1 closely follows that of Theorem 1 of Holm, Mark and Adolfsson [5], but our result enjoys advantages of generality, simplicity, and rigor. Concerning generality, their result is for a specific method of analysis of absolute normals, whereas the theorem here applies to a wide class of methods, as will be demonstrated in Section 3, and under relaxed distributional assumptions. Concerning simplicity, their method focuses and explicitly relies on the use of coverage bounds for values in an ordered random sample, whereas the method here does not—a conceptual simplification that greatly simplifies implementation. In addition, we have made the proof rigorous.

Theorem 1 very much depends on the independence of the estimators, since Lemma 1 does.

Theorem 1 also hinges very much on theorem conditions (i) and (ii). By Corollary 1, these conditions hold for absolute or squared normal estimates, scaled to have common variance. It is natural to ask, for what other families $f(x;\theta)$ of densities with positive support do conditions (i) and (ii) hold. The following lemmas shed some light on this question.

Lemma 4 *Consider $f(x;\theta)$, $x > 0$. If $f(x;0)/f(x;\theta)$ is non-increasing in x for each $\theta \neq 0$, then $X_{0c} \preceq X_{\theta c}$ for any constant $c > 0$.*

Proof. For $X \sim f(x)$ with cdf $F(x)$, and for any constant $c > 0$, $X_c \sim cf(cx)/F(c)$, $x > 0$. Since $f(x;0)/f(x;\theta)$ is non-increasing in x for each $\theta \neq 0$, it follows that $f_c(x;0)/f_c(x;\theta)$ is non-increasing in x for each $\theta \neq 0$, for f_c the density of X_c. The result follows. □

So, a monotone likelihood ratio property for the family $f(x;\theta)$ is sufficient for (i) of Theorem 1. Unfortunately, as will be seen next, the corresponding sufficient condition for (ii) is a density ratio property of the null distribution, not of the family.

Lemma 5 *Consider $X \sim f(x)$, $x > 0$. Any of the followings is sufficient for $X_{0d} \preceq X_{0c}$ for any $d > c > 0$.*

(i) $f(dx)/f(cx)$ is non-increasing in x for any $d > c > 0$;

(ii) $cf'(cx)/f(cx)$ is non-increasing in c for all x.

Proof. Let $f(x)$ and $F(x)$ denote the pdf and cdf of X, respectively, and $f_c(x)$ the pdf of X_{0c}. Then $f_d(x)/f_c(x) = [dF(c)/(cF(d)][f(dx)/f(cx)]$. Hence, $f(dx)/f(cx)$ nonincreasing in x implies $f_d(x)/f_c(x)$ nonincreasing in x, so $X_{0d} \preceq X_{0c}$, establishing (i).

To prove (ii), consider $r(x) = f(dx)/f(cx)$ for any $d > c > 0$. Then $r(x)$ is nonincreasing for all $d > c > 0$ if $r'(x) \leq 0$ which follows if $cf'(cx)/f(cx)$ is nonincreasing in c for all x. □

The following characterizes property (ii) of the theorem.

Lemma 6 *Consider $X \sim f(x)$, $x > 0$. Then $X_{0d} \preceq X_{0c}$ for any $d > c > 0$ if and only if $F(cx)/F(c)$ is nondecreasing in c for all $0 < x < 1$.*

Proof. The cdf of X_c is $F_c(x) = F(cx)/F(c)$, for $0 < x < 1$. □

In the next section, several step-down tests proposed in the literature, variations on such tests, and iterative variations on other tests, are shown to strongly control the familywise error rate as a consequence of Theorem 1.

3 Applications to specific iterative step-down tests

Many step-down procedures have been proposed in the literature for the analysis of saturated or nearly saturated orthogonal designs, without a proof that they provide strong control of the familywise error rate. Many of these, or applicable variations on these, do provide strong control of the familywise error rate by virtue of Theorem 1, as noted in this section. To simplify the presentation in this section, consider normally distributed estimators $\hat{\theta}_i \sim N(\theta_i, \sigma^2)$.

3.1 Tests of Zahn [19, 20] and Venter and Steel [11]

Iterative step-down methods based on either absolute estimates $X_i = |\hat{\theta}_i|$ or squared estimates $X_i = \hat{\theta}_i^2$ were considered by Zahn [19, 20] and Venter and Steel [11]. Tests of Daniel [4] and Birnbaum [3] for the effect with largest estimate, if applied iteratively, are special cases of the tests of Zahn [19, 20]. Each of these methods uses test statistics of the form

$$T_i = X_{[i]} / \left[\sum_{h=1}^{i} a_{hi} X_{[h]} \right],$$

for prespecified nonnegative constants a_{hi} not all zero. In each case, the critical value c_i for T_i is obtained as the upper-α quantile of the null distribution of T_i as a function of only i independent folded standard normal or χ_1^2 random variables $Y_1, \ldots, Y_i$. The folded normal distribution is used for analysis of the absolute estimates $X_i = |\hat{\theta}_i|$ and the Chi-squared distribution for analysis of the squared estimates $X_i = \hat{\theta}_i^2$.

Corollary 2 *The iterative step-down tests of Zahn* [19, 20] *and Venter and Steel* [11] *each provides strong control of the familywise error rate.*

3.2 Tests of Voss [12] and Voss and Wang [14]

Voss [12] proposed a closed step-down procedure that simultaneously tests the hypotheses $H_{0i} : \theta_i = 0$, $i = 1, 2, \ldots, k$, using for example the test statistics

$$T_i = X_{[i]} / \left[\sum_{h=1}^{\nu} X_{[h]} / \nu \right] \tag{2}$$

for fixed $\nu < k$, using either absolute estimates $X_i = |\hat{\theta}_i|$ or squared estimates $X_i = \hat{\theta}_i^2$. Critical values for a closed step-down test depend on the distribution of all k estimators whereas, as noted by Voss [12], the corresponding iterative test uses sharper critical values c_i that depend on the distribution of only i estimators. In particular, consider the iterative step-down test using test statistics

$$T_i = X_{[i]} / \left[\sum_{h=1}^{\nu_i} X_{[h]} / \nu_i \right] \tag{3}$$

for prespecified integers $\nu_i \leq i$ for $i = 2, \ldots, k$, where the critical value c_i for T_i is obtained as the upper-α quantile of the distribution of $T_i = Y_{(i)} / \left[\sum_{h=1}^{\nu_i} Y_{(h)} / \nu_i \right]$ as a function of independent folden standard normal or χ_1^2 random variables $Y_1, \ldots, Y_i$ as appropriate.

Voss and Wang [14] generalized the test of Voss [12] to include adaptive test statistics. In particular, for each test statistic T_i in equation (2), the denominator $\left[\sum_{h=1}^{\nu} X_{[h]} / \nu \right]$ could be replaced by the minimum of multiple functions $a_j \left[\sum_{h=1}^{\nu_j} X_{[h]} / \nu_j \right]$

indexed by j, for prespecified constants a_j and integers ν_j. Voss and Wang [14] showed that familywise error rate is still strongly controlled in this case. As a variation on this, consider the iterative, adaptive, step-down test obtained by replacing the denominator of T_i in equation (3) by the minimum of multiple functions of the form $a_{ij}\left[\sum_{h=1}^{\nu_{ij}} X_{[h]}/\nu_{ij}\right]$ indexed by j, for prespecified constants $a_{ij} > 0$ and integers $\nu_{ij} \leq i$.

Corollary 3 *The above iterative variations on the step-down tests of Voss* [12] *and Voss and Wang* [14] *provide strong control of the familywise error rate.*

3.3 Langsrud and Naes' [6] test for nearly saturated designs

Langsrud and Naes [6] considered a similar iterative step-down test based on the squared estimates $X_i = \hat{\theta}_i^2$ but incorporating an independent estimator $\hat{\sigma}^2 = SSE/\nu$ based on ν error degrees of freedom. Their method uses test statistics of the form

$$T_i = X_{[i]}/\left[\left(\sum_{h=1}^{\nu_i} X_{[h]} + SSE\right)/(\nu_i + \nu)\right],$$

for prespecified value of $\nu_i < i$. Introducing the independent variance estimator in the denominator of each test statistic introduces no technical difficulties with respect to stochastic ordering. The critical value c_i for T_i is obtained as the upper-α quantile of the null distribution of T_i as a function of i independent χ_1^2 random variables $Y_1, \ldots, Y_i$ and an independent variance estimator $\hat{\sigma}^2$, where $\nu\hat{\sigma}^2/\sigma^2 \sim \chi_\nu^2$.

Corollary 4 *The iterative step-down test of Langsrud and Naes* [6] *provides strong control of the familywise error rate.*

3.4 A generalization of Holm, Mark and Adolfsson's [5] test

In this paper we generalize the result of Holm, Mark and Adolfsson [5], who provided an adaptive step-down test based on the absolute estimates $X_i = |\hat{\theta}_i|$. As we interpret it, their method in essence uses test statistics of the form

$$T_i = X_{[i]}/\left[\min_{h \leq i} a_{hi} X_{[h]}\right], \tag{4}$$

for prespecified constants $a_{hi} > 0$; $h = 1, \ldots, i$; $i = 1, \ldots, k$. The constants a_{hi} are implicitly determined in their method when they obtain coverage bounds for order statistics of a random sample. Each $g_{hi}(X_{[1]}, \ldots, X_{[i]}) = a_{hi}X_{[i]}$ in the denominator of T_i in equation (4) is an exchangable, nondecreasing function of $(X_{[1]}, \ldots, X_{[i]})$, and so is the minimum $g_i(X_{[1]}, \ldots, X_{[i]}) = \min_{h \le i} g_{hi}(X_{[1]}, \ldots, X_{[i]})$ of such functions, used as the denominator of T_i. They choose the constants a_{hi} based on consideration of one-sided coverage bounds for estimates, but that is beside the point here.

Using $g_i(X_{[1]}, \ldots, X_{[i]}) = \min_{h \le i} a_{hi}X_{[h]}$ makes the method adaptive, since which order statistic $X_{(h)} = X_{[h]}$ yields the denominator $a_{hi}X_{[h]}$ of T_i is random, depending on the data. Wang and Voss [16] and Voss and Wang [14] made similar use of the minimum function to establish adaptive confidence intervals and step-down tests, respectively, providing strong control of familywise error rates. By this approach, one can generalize the method of Holm, Mark and Adolfsson [5] as follows. Consider as test statistics

$$T_i = X_{[i]} / \left[\min_{h \in H_i} g_{hi}(X_{[1]}, \ldots, X_{[i]}) \right] , \tag{5}$$

where H_i is the index set for a collection of functions for each i, and each g_{hi} is of the form $g_{hi}(X_{[1]}, \ldots, X_{[i]}) = \sum_{j=1}^{i} a_{hij}X_{[j]}$ for nonnegative constants a_{hij} not all zero for each hi. This extension allows the denominator to involve linear combinations of the $X_{[j]}$ rather than simply a multiple of one of these.

In each of these variations, the critical value c_i for T_i is obtained as the upper-α quantile of the null distribution of T_i as a function of i independent folded standard normal random variables $Y_1, \ldots, Y_i$. Alternatively, one could use the squared estimators $X_i = \hat{\theta}_i^2$ to compute the test statistics, in which case the critical value c_i would be computed from independent χ_1^2 random variables $Y_1, \ldots, Y_i$.

Corollary 5 *The generalization of the iterative step-down test of Holm, Mark and Adolfsson* [5] *in equation (5) provides strong control of the familywise error rate.*

3.5 Iterative Lenth-like tests

Lenth [8] provided a “quick and easy” adaptive method of analysis of unreplicated factorials that has spurred great interest in adaptive methods of analysis of such designs. Using the absolute estimates $X_i = |\hat{\theta}_i|$, Lenth's method is as follows. First compute an initial estimate of σ as $\hat{\sigma}_0 = 1.5 \times \text{median}\,\{X_1, \ldots, X_k\}$. Then obtain a second, *adaptive* estimate as $\hat{\sigma} = 1.5 \times \text{median}\,\{X_i : X_i \le 2.5\hat{\sigma}_0\}$. Approximate inferences are based on $(\hat{\theta}_i - \theta_i)/\hat{\sigma}$. Ye, Hamada and Wu [18] proposed applying

Lenth's [8] method stepping down iteratively. It remains an open problem to show that either Lenth's [8] method or the variation of Ye, Hamada and Wu [18] provides strong control of the familywise error rate, because of the non-monotone way in which their denominators are adaptive. However, consider applying the iterative step-down test of Section 2 as follows.

Use as test statistics

$$T_i = X_{[i]} / \left[1.5 \times \text{median}\left\{ X_{[1]}, \ldots, X_{[i]} \right\} \right].$$

Obtain the critical value c_i for T_i as the upper-α quantile of the null distribution of T_i as a function of i independent folded standard normal random variables $Y_1, \ldots, Y_i$.

Corollary 6 *The above iterative step-down test, a variation on Lenth's* [8] *adaptive test, provides strong control of the familywise error rate.*

One would anticipate this variation to have similar operating characteristics to the methods of Lenth [8] and Ye, Hamada and Wu [18], but with the additional guarantee that the familywise error rate is strongly controlled. One could say that iterating a non-adaptive method makes it adaptive, since any very large effects are highly likely to be asserted nonzero, in which case they would have no impact on the analysis of the other effects.

4 Closing remarks

Recently Holm, Mark and Adolfsson [5] provided the first iterative step-down test for analysis of orthogonal saturated designs with a proof that the familywise error rate is strongly controlled over all parameter configurations. Using an essential aspect of their technical approach, we have established strong control of the familywise error rate for a large class of iterative step-down tests, including a generalization of the Holm, Mark and Adolfsson [5] test, a number of other iterative step-down tests proposed in the literature, and variations on other multiple tests proposed in the literature. While the results presented here generalize those of Holm, Mark and Adolfsson [5] to a much broader class of test statistics, this is achieved without the explicit consideration of coverage bounds that seems fundamental to their methodology, providing a simplification of both the theory and the implementation. The results also apply to certain step-down tests for the analysis of nearly saturated designs. Furthermore, the usual normality assumption has been relaxed.

Acknowledgments

The authors are grateful for the efforts and constructive suggestions of two referees and an associate editor.

References

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

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

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

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

[5] Holm, S., Mark, S., and Adolfsson, T. (2005). A step-down test for effects in unreplicated factorial designs. *Comm. Statist. Theory Methods* **34**, 405–416.

[6] Langsrud, O. and Naes, T. (1998). A unified framework for significance testing in fractional factorials, *Comput. Statist. Data Analysis* **28**, 413–431.

[7] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

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

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

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

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

[12] 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.

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

[14] Voss, D. T. and Wang, W. (2006). On adaptive testing in orthogonal saturated designs. *Statist. Sinica*, **16**, 227–234.

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

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

[17] Wu, S. S. and Wang, W. (2006). Step-up simultaneous tests for identifying active effects in orthogonal saturated designs. To appear in *Ann. Statist.*

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

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

[20] 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, OHIO 45435-0001, USA
E-MAIL: dan.voss@wright.edu
E-MAIL: wwang@euler.wright.edu