SCIENCE CHINA
Mathematics

• ARTICLES •

April 2010 Vol. 53 No. 4: 927–938
doi: 10.1007/s11425-010-0071-1

On hypothesis testing with a partitioned random alternative

WANG WeiZhen

Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435*, USA*
Email: weizhen.wang@wright.edu

Received March 1, 2008; accepted June 16, 2009

Abstract It is common in statistical practice that one needs to make a choice among $m+1$ mutually exclusive claims on distributions. When $m=1$, it is done by the (traditional) hypothesis test. In this paper, a generalization to the case $m>1$ is proposed. The fundamental difference with the case $m=1$ is that the new alternative hypothesis is a partition of m multiple claims and is data-dependent. Data is used to decide which claim in the partition is to be tested as the alternative. Thus, a random alternative is involved. The conditional and overall type I errors of the proposed test are controlled at a given level, and this test can be used as a new solution for the general multiple test problem. Several classical problems, including the one-sample problem, model selection in multiple linear regression, and multi-factor analysis, are revisited, and new tests are provided correspondingly. Consequently, the famous two-sided t-test should be replaced by the proposed.

Keywords analysis of variance, conditional probability, multiple linear regression, one-sample problem

MSC(2000): 62F03, 62A01

Citation: Wang W Z. On hypothesis testing with a partitioned random alternative. Sci China Math, 2010, 53(4): 927–938, doi: 10.1007/s11425-010-0071-1

1 Introduction

The idea of the (traditional) statistical hypothesis test is to verify what we believe on the distribution involved in the study. The null and alternative hypotheses are statements regarding the distribution or parameters, and should be independent of the data. This approach is efficient if we understand the situation well, i.e., a small alternative hypothesis is formulated. This is because establishing the alternative hypothesis is highly reliable and a small alternative provides precise information for the location of parameters. In many cases, however, the alternative hypothesis space is very large. Establishing such a hypothesis does not provide much useful information about parameters to the practitioner. The situation gets worse when the study involves multiple parameters, as shown in the following two examples.

Example 1 (One sample problem with a two-sided alternative)**.** Consider

$$H_0: \mu = \mu_0 \text{ vs } H_A: \mu \neq \mu_0 \tag{1.1}$$

for a given value μ_0 based on an i.i.d. sample $\{X_i\}_{i=1}^n$ from a normal population $N(\mu,\sigma^2)$ with unknown mean μ and variance σ^2. Let $\bar{X}$ and S denote the sample mean and standard deviation for the sample. The famous two-sided t-test establishes H_A at level α if

$$|T_{\text{obs}}| \stackrel{\text{def}}{=} \left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right| > t_{\alpha/2;n-1}, \tag{1.2}$$

where $t_{\alpha/2;n-1}$ is the upper $\alpha/2$th percentile of a t-distribution with $n-1$ degrees of freedom. However, it cannot establish $\mu > \mu_0$ if $T_{\rm obs} > t_{\alpha/2;n-1}$ unless the one-sided hypotheses

$$H_0 : \mu = \mu_0 \ \text{ vs } \ H_A : \mu > \mu_0 \tag{1.3}$$

were tested in the first place, though $T_{\rm obs} > t_{\alpha/2;n-1} > t_{\alpha;n-1}$. A natural question is then raised: can we pick an appropriate alternative for testing based on the data? Intuitively, if $T_{\rm obs} > 0$, one should test for (1.3); if $T_{\rm obs} < 0$, then test for

$$H_0 : \mu = \mu_0 \ \text{ vs } \ H_A : \mu < \mu_0; \tag{1.4}$$

and ignore the case of $T_{\rm obs} = 0$ because the last has a zero probability under any value of parameter μ. Clearly, the alternative is data-dependent.

Example 2 (Model selection in multiple linear regression)**.** A data set presented in [13, p. 303] concerns the heat evolved of cement (the response Y) as a linear function of four possible predictors: tricalcium aluminate (X_1), tricalcium silicate (X_2), tetracalcium aluminoferrite (X_3) and dicalcium silicate (X_4). Consider the standard linear regression model (4.1) with $n = 13$ observations and $p = 4$. If testing

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \ \text{ vs } \ H_A : \text{not all } \beta_j\text{'s equal zero},$$

the overall F-test (=MSR/MSE) leads to an establishment of H_A at level 0.05. However, which predictors should be included in or excluded from the regression model cannot be answered by establishing the alternative.

In general, suppose a random vector X is observed from a sample space S with a probability distribution P_θ and the parameter vector θ belongs to a parameter space Θ with a partition $\Theta = \Theta_0 \cup \Theta_A$. The basic hypotheses are

$$H_0 : \theta \in \Theta_0 \ \text{ vs } \ H_A : \theta \in \Theta_A. \tag{1.5}$$

The alternative Θ_A, for example in Example 2, is too large to locate θ precisely. Let $\{\Theta_{iA}\}_{i=1}^m$ be a partition of Θ_A, where each Θ_{iA} represents an interesting case of parameter. Correspondingly, on the sample space S, let $\{S_i\}_{i=1}^m$ be m mutually exclusive sets with

$$P_\theta\left(\bigcup_{i=1}^m S_i\right) = 1, \quad \forall\ \theta \in \Theta. \tag{1.6}$$

We call $\{S_i\}_{i=1}^m$ an almost surely partition of S. The set S_i should be more in favor of Θ_{iA} than any other Θ_{jA}. On each S_i, a choice between Θ_0 and Θ_{iA} is made using the traditional statistical test. The hypotheses of interest for this paper are

$$H_0 : \theta \in \Theta_0 \ \text{ vs } \ H_A : \theta \in \Theta_{iA}, \quad \text{if } X \in S_i, \text{ for } i = 1, \ldots, m, \tag{1.7}$$

or equivalently

$$H_0 : \theta \in \Theta_0 \ \text{ vs } \ H_A : \theta \in \bigcup_{i=1}^m (\Theta_{iA} I_{S_i}(x)), \tag{1.8}$$

where $I_{S_i}(\cdot)$ is an indicator function, and $\Theta_{iA}0 \stackrel{\rm def}{=} \emptyset$ (the empty set) and $\Theta_{iA}1 \stackrel{\rm def}{=} \Theta_{iA}$. For example, if $X \in S_2$, then we test Θ_0 against Θ_{2A} *but using S_2, instead of S, as the sample space.*

Unlike in the traditional test, now the alternative, as a mapping from S to Θ_A, is NOT constant-valued. It equals Θ_{iA} on S_i. So a random alternative is formulated. We first use data to determine which alternative Θ_{iA} should be tested by checking which S_i includes the data, then decide whether Θ_{iA} is accepted using certain test statistic(s). It is a two-stage data analysis. When $m = 1$, however, (1.7) reduces to the traditional hypotheses.

Section 2 proposes a method to construct level-α tests for (1.7). In Sections 3, 4 and 5, we discuss applications of the proposed method to three classical problems: the one-sample problem, multiple linear regression and multi-factor analysis. More precisely, we show in Theorem 3.1 in Section 3 that the traditional one sample two-sided t-test (1.2) should be replaced by the proposed due to the less precise alternative hypothesis of the former. In Section 4, the goal is to identify all nonzero coefficients in a multiple linear regression equation, i.e., model selection. Note that even though there exist several step-related methods to select a model in multiple linear regression, none of them is able to control any reasonable error rate. In fact, researchers have still not even reached an agreement on the kind of error to be controlled. But for all cases discussed in this paper, the type I errors are always controlled at the nominal level. In Section 5, we detect factor main effects and interactions in one step. A single test is valid for all three two-way anova models with fixed, or mixed or random effects. Also for three-way anova with random effects, the approximate F-tests (the Satterthwaite's type) are no longer needed. Discussion is given in Section 6.

2 A new testing procedure with a random alternative

When two partitions $\{\Theta_{iA}\}_{i=1}^m$ and $\{S_i\}_{i=1}^m$ are available, one is ready to derive a test for (1.7). The union of S_i's can be used as a reduced sample space since its total probability is equal to one. On each S_i, conduct a level α test for

$$H_0 : \theta \in \Theta_0 \text{ vs } H_A : \theta \in \Theta_{iA} \tag{2.1}$$

with S_i as the sample space. Let R_i (a subset of S_i) be a level-α acceptance region of Θ_{iA}, i.e.,

$$\sup_{\theta\in\Theta_0} P_\theta(R_i|S_i) \leqslant \alpha. \tag{2.2}$$

Then the overall rejection region of H_0 for (1.7) is given by

$$R = \bigcup_{i=1}^m R_i. \tag{2.3}$$

The overall type I error is defined as

$$P_\theta(R) \text{ for } \theta \in \Theta_0, \tag{2.4}$$

and each conditional type I error (for a fixed $i = 1, \dots, m$) is given by

$$P_\theta(R_i|S_i) \text{ for } \theta \in \Theta_0, \tag{2.5}$$

which is controlled at level α by (2.2). Furthermore, the conditional power of the test corresponding to R_i is defined as

$$\text{Power}_{R_i}(\theta) = P_\theta(R_i|S_i) = \frac{P_\theta(R_i)}{P_\theta(S_i)}, \quad \forall\, \theta \in \Theta_{iA}. \tag{2.6}$$

Theorem 2.1. *The rejection region R given in* (2.3) *defines a level α test for* (1.7)*, as well as for* (1.5).

Proof. Recall $R_i = R \cap S_i$. For any $\theta \in \Theta_0$,

$$P_\theta(R) = \sum_{i=1}^m P_\theta(R \cap S_i) = \sum_{i=1}^m P_\theta(R_i|S_i)P_\theta(S_i) \leqslant \alpha \sum_{i=1}^m P_\theta(S_i) = \alpha.$$

Therefore, the test size is no larger than α.

Remark 1. The concept of the rejection region R (to reject the null hypothesis) is not clear enough to specify all $m+1$ possible actions taken in the new procedure: declare Θ_{iA} for $i = 1, \dots, m$ and Θ_0. The almost surely partition $\{S_i\}_{i=1}^m$ is also needed. For example, if $R_i\,(= R \bigcap S_i)$ occurs for some i, then accept Θ_{iA}; otherwise, accept Θ_0. Therefore, a notation of "$R\,(= \bigcup_{i=1}^m R_i)$" is preferred to "$R$" itself, which reminds one that the actions taken on $R\,(= \bigcup_{i=1}^m R_i)$ vary and depend on R_i. So from now on we call R_i the acceptance region of Θ_{iA}, but still call R the rejection region of Θ_0.

There are three levels of using data in hypothesis tests. The first is to determine a rejection region of H_0 as in the traditional test. The second is to partition the sample space and use a different test statistic in each set of the partition to construct a rejection region. This was originally proposed by Hogg, Fisher and Randles [7]. See [14] for more results along this research line. The third is the proposed here that generates two partitions: one in the sample space and one in the alternative, and constructs a level-α acceptance region of Θ_{iA} in each set of the former partition. To see the difference between the second and the third, Hogg, Fisher and Randles [7] define a region $R_i \subset S$, instead of $R_i \subset S_i$ as we propose, for (2.1) satisfying

$$\sup_{\theta \in \Theta_0} P_\theta(R_i) \leqslant \alpha \tag{2.7}$$

instead of (2.2). As a result, they need a critical condition — independence of S_i and R_i — to show their type I error to be controlled at level α. Our proposed test, however, controls the type I error automatically, and thus is much more general and flexible. In fact, Hogg and Lenth [8, p. 1562] gave a simple example of Theorem 2.1 for the two sample t-test problem. However, they were unable to obtain more, and they wrote: "It would be a challenging problem to extend the above idea and develop an adaptive order-related test of several means as an alternative to the analysis of variance and the associated multiple-comparison procedures."

Typically, based on an understanding of the purpose of a data analysis, $\{\Theta_{iA}\}_{i=1}^m$ can easily be obtained. The choice on partition $\{S_i\}_{i=1}^m$, however, is critical. Two methods to obtain this partition are introduced below. The first is simple but has fewer applications, and the second is general but may require extensive numerical computation.

I) When Θ_{iA}, for $i \geqslant 1$, are disjoint open sets in Euclidean space, there typically exists a natural choice on S_i, as shown in (3.1). Another example is to identify the single largest treatment mean in the one-way anova, which deserves a complete discussion in a different paper.

II) In this second method, we first compute a p-value for each set of hypotheses (2.1) based on a certain test statistic T_i with a realization t_i, denoted by $p_i(t_i)$, using the original sample space S as the sample space. For an given observed data set, since the distributions of T_i's are typically different, instead of comparing all t_i's, we compare all p-values $p_i(t_i)$'s. Note two facts: 1) the null hypothesis in (2.1) is the same for all i's, and 2) the sample space is always equal to S. Therefore, the Θ_{iA} with the smallest p-value is most likely to be true, which results in a choice of S_i,

$$S_i \stackrel{\text{def}}{=} \{x \in S : p_i(t_i) < \min\{p_j(t_j) : j \neq i\}\}, \tag{2.8}$$

provided that any two p-values have zero probability to be equal to each other. In short, S_i consists of those sample points on which Θ_{iA} is most likely to be true. This construction of S_i's is general and seems objective due to 1) and 2), and so is recommended in this paper. Also, if checking (3.1) carefully, the S_i obtained by Method I) is equal to the one generated by Method II). However, Method II) may require extensive computation of p-values especially when m is large. It is applied in Sections 4 and 5, and can also be applied to the problems of identifying active effects in orthogonal or nonorthogonal saturated designs and many others.

There have been many efforts to resolve the multiple test problem. In simultaneous testing, several pairs of hypothesis are formulated and the data set plays no role in determining which pairs of hypotheses should be tested. So all pairs of hypothesis have to be tested. In consequence, what kind of error to be controlled has to be determined. Controlling individual type I error is not good enough, but controlling the experimentwise error rate (EWER) is a very challenging task, and may result in conservative tests. The simplest way to control the EWER is to use Bonferroni inequality on the levels of all individual tests. However, the resultant tests are very conservative. Improvements are given by Holm [9] and Hochberg [6]. Marcus, Peritz and Gabriel [12] proposed a closed test procedure with a controlled EWER. Dunnett and Tamhane [3] discussed a step-up test procedure, in which the hypotheses tested are indeed data-dependent, though it is not clearly stated. This is also true for the step-down test procedure. Benjamini and Hochberg [2] proposed to control the false discovery rate for multiple tests. Finner and Strassburger [4] used a partition on the parameter space to conduct tests. Lehmann and Romano [11] proposed to control the error of making at least $k\,(\geqslant 1)$ incorrect assertions, which is a generalization of the EWER (i.e., $k = 1$). The motivation of all these efforts is to seek an appropriate balance between controlling a certain type of error and obtaining powerful tests. But the implementation is still in question.

Compared with these methods, our proposed utilizes two related partitions, one in the alternative space and one in the sample space, to convert a multiple test problem to a single test problem. Thus only the type I error needs to be controlled. In fact, we control all conditional type I errors. On each set in the partition of the sample space, a traditional test is conducted, so the implementation is relatively simple.

In summary, the proposed test consists of:

a) a partition $\{\Theta_{iA}\}_{i=1}^{m}$ on Θ_A;

b) an almost surely partition $\{S_i\}_{i=1}^{m}$ on S;

c) an acceptance region R_i in each S_i satisfying $\sup_{\theta\in\Theta_0} P_\theta(R_i|S_i) = \alpha$.

3 Application in the one-sample problem

The one-sample problem (1.1) is one of the basic statistical problems and is widely used in practice. We derive a test for this problem by applying Theorem 2.1. A power comparison with the traditional t-test (1.2) is provided in Theorem 3.1.

Example 1 (continued). The alternative hypothesis space $\Theta_A = \{\mu \neq \mu_0\}$ is equal to the union of two disjoint open sets

$$\Theta_{1A} = \{\mu > \mu_0\} \text{ and } \Theta_{2A} = \{\mu < \mu_0\}$$

in R^1. Thus, following Method I), there exists a natural choice of the partition on S:

$$S_1 = \{\bar{X} > \mu_0\} \text{ and } S_2 = \{\bar{X} < \mu_0\}. \tag{3.1}$$

Following Theorem 2.1, on S_1, conduct a test of (1.3) with acceptance region of Θ_{1A}

$$R_1 = \{T_{\rm obs} > t_{\alpha/2;n-1}\} \text{ satisfying } P_{\mu=\mu_0}(R_1|S_1) = \alpha.$$

Similarly, on $S_2 = \{\bar{X} < \mu_0\}$, conduct a test of (1.4) with acceptance region of Θ_{2A}

$$R_2 = \{T_{\rm obs} < -t_{\alpha/2;n-1}\} \text{ satisfying } P_{\mu=\mu_0}(R_2|S_2) = \alpha.$$

Therefore, with an overall rejection region

$$R\,(= R_1 \cup R_2), \tag{3.2}$$

one accepts $\mu > \mu_0$ on R_1 and $\mu < \mu_0$ on R_2, respectively.

Remark 2. Even though test (3.2) has the same rejection region (as a set) as the traditional two-sided t-test (1.2), the two tests are different since the former takes two actions: declare $\mu < \mu_0$ or $\mu > \mu_0$ on $R(= R_1 \cup R_2)$ depending on R_i, while the latter only takes a single action: declare $\mu \neq \mu_0$ on R.

Theorem 3.1. *Given each S_i, the probability of detecting Θ_{iA} for* (3.2) *is equal to that of detecting $\mu \neq \mu_0$ for the t-test* (1.2) *for any $\mu \in \Theta_{iA}$, i.e.,*

$$\text{Power}_{R_i}(\mu,\sigma) = P_{\mu,\sigma}(R|S_i), \quad \textit{for } \mu \in \Theta_{iA}. \tag{3.3}$$

Proof. The claim in (3.3) is trivial because $R \cap S_i = R_i$.

Remark 3. An analogous result can be established for any testing problem with a two-sided alternative hypothesis and a t-test statistic, including the two-independent samples problem (see also [8]), the match paired t-test problem and the test for one coefficient in a simple or multiple linear regression. Due to (3.3), the corresponding t-tests should be replaced by the proposed because they have the same conditional power as the proposed, but draw less precise conclusions on the parameter than the proposed.

4 Application in multiple linear regression

Suppose a response variable Y and a group of potential predictors $\{X_j\}_{j=1}^{p}$ are observed from a subject. It is of great interest to determine which predictors should be included in the regression model. Many methods have been proposed on this issue. The most commonly used include forward selection, backward elimination, stepwise and all subsets methods. See, for example, Section 9.2 in [13] for details. It is interesting to notice that the step-related procedures also use the data to determine which predictors should be included or excluded from the model at each step, i.e., testing whether the corresponding coefficient in the regression model is equal to zero. However, they do not have a well-defined partition on the alternative hypothesis. More importantly, none of them control the error rate even though they are all based on hypothesis tests. In fact, the researchers do not even have a common agreement about what error should be controlled.

Assume the standard multiple linear regression model (the full model):

$$Y_i = \mu + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i, \ \text{ for } \ i = 1, \ldots, n\,(> p+1), \tag{4.1}$$

where $\epsilon_i \sim$ i.i.d. $N(0, \sigma^2)$. The unknown parameters β_j's are of interest and μ and σ are two unknown nuisance parameters. Assume that the design matrix ($= \{X_{ij}\}$ for $1 \leqslant i \leqslant n$ and $0 \leqslant j \leqslant p$ with $X_{i0} = 1$) is of full rank of $p+1$. The goal here is to identify all nonzero β_j's. The parameter vector is $\theta = (\mu, \beta_1, \ldots, \beta_p, \sigma^2)$. Let $\theta_0 = (0, 0, \ldots, 0, 1)$. The basic hypotheses are

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0 \ \text{ vs } \ H_A: \ \text{ not all } \beta_j \text{ equal zero } (\stackrel{\text{def}}{=} \Theta_A^{LR}), \tag{4.2}$$

where the alternative cannot identify nonzero β_j and the superscript "LR" stands for linear regression.

First obtain a partition of H_A in (4.2). For any $\theta \in \Theta_A^{LR}$, let N_R be the number of β_j's being zero. The possible values of N_R are 0 through $p-1$ (note that N_R cannot equal p in H_A and $N_R = p-1$ means a simple linear regression). Thus, the alternative space can be partitioned as

$$\Theta_A^{LR} = \bigcup_{j=0}^{p-1} \Theta_{jA}^{LR} = \bigcup_{j=0}^{p-1} \bigcup_{\text{all } I_j} \Theta_{I_j A}^{LR}, \tag{4.3}$$

where

$$\Theta_{jA}^{LR} = \{N_R = j\}, \ \ \Theta_{I_j A}^{LR} = \{\beta_u = 0 \ \text{ if and only if } \ u \in I_j\},$$

and I_j is a subset of $I = \{1, \ldots, p\}$ with j elements. When $j = 0$, I_j is the empty set. There is no further partition on $\Theta_{\emptyset A}^{LR}$ and

$$\Theta_{\emptyset A}^{LR} = \{N_R = 0\} = \{\beta_u \neq 0, \forall \ 1 \leqslant u \leqslant p\}.$$

Note that each $\Theta_{I_j A}^{LR}$ contains the exact information needed for the β_j, and $\{\Theta_{I_j A}^{LR}\}_{\text{all } I_j,\ 0 \leqslant j \leqslant p-1}$ is the partition on Θ_A^{LR}.

Secondly, we provide results needed to construct a partition of the sample space using Method II). Let $\{\hat\beta_j\}_{j=1}^{p}$ be the least squares estimators under model (4.1) with covariance matrix $\sigma^2\Sigma$ (note that Σ is a known matrix). For a given set I_j with $j \in [0, p-1]$, let $\hat\beta_{I_j} = \{\hat\beta_u\}_{u \in I_j}$ and let $\sigma^2\Sigma_{I_j}$ be the covariance matrix of $\hat\beta_{I_j}$. Also let

$$F_{I_j}^{LR} = \frac{(\hat\beta_I \Sigma_I^{-1} \hat\beta_I' - \hat\beta_{I_j} \Sigma_{I_j}^{-1} \hat\beta_{I_j}')/(p-j)}{(\hat\beta_{I_j} \Sigma_{I_j}^{-1} \hat\beta_{I_j}' + SSE)/(n-p-1+j)}, \tag{4.4}$$

and let $f_{I_j}^{LR}$ be a realization of $F_{I_j}^{LR}$.

Lemma 4.1. *Under H_0 in (4.2), $F_{I_j}^{LR}$ follows an F-distribution with $p-j$ and $n-p-1+j$ degrees of freedom. For each observed $f_{I_j}^{LR}$, the random variable*

$$p_{I_j}(f_{I_j}^{LR}) \stackrel{\text{def}}{=} P_{\theta_0}(F_{I_j}^{LR} \geqslant f_{I_j}^{LR}) = P(F_{p-j, n-p-1+j} \geqslant f_{I_j}^{LR}) \tag{4.5}$$

defines a p-value for the hypotheses:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0 \ \ vs \ \ H_A : \ \ \theta \in \Theta^{LR}_{I_jA}. \tag{4.6}$$

Proof. Write

$$\hat\beta_I = (\hat\beta_{I-I_j}, \hat\beta_{I_j}), \quad \Sigma_I = \begin{pmatrix} A & B \\ B' & \Sigma_{I_j} \end{pmatrix}.$$

Following Corollary A.3.2 in [1],

$$\hat\beta_I \Sigma_I^{-1} \hat\beta'_I - \hat\beta_{I_j} \Sigma_{I_j}^{-1} \hat\beta'_{I_j} = (\hat\beta_{I-I_j} - B\Sigma_{I_j}^{-1}\hat\beta_{I_j}) G^{-1} (\hat\beta_{I-I_j} - B\Sigma_{I_j}^{-1}\hat\beta_{I_j})', \tag{4.7}$$

where $G = A - B\Sigma_{I_j}^{-1}B'$. Note that under H_0 the conditional distribution of $\hat\beta_{I-I_j}$ given $\hat\beta_{I_j}$ is normal with mean vector $B\Sigma_{I_j}^{-1}\hat\beta_{I_j}$ and covariance matrix G. See, for example, Theorem 2.5.1 in [1]. We conclude the independence between $\hat\beta_{I-I_j} - B\Sigma_{I_j}^{-1}\hat\beta_{I_j}$ and $\hat\beta_{I_j}$ and that $F^{LR}_{I_j}$ follows an F-distribution. Therefore, a level-α test for (4.6) can be obtained by rejecting H_0 if $p(f^{LR}_{I_j})$ is no larger than α.

Consider any fixed j. For any two subsets I_j and I'_j, since $F^{LR}_{I_j}$ and $F^{LR}_{I'_j}$ have the same distribution under H_0, it is clear that $p_{I_j}(f^{LR}_{I_j}) < p_{I'_j}(f^{LR}_{I'_j})$ if and only if $f^{LR}_{I_j} > f^{LR}_{I'_j}$. Thus between two sub-alternatives $\Theta^{LR}_{I_jA}$ and $\Theta^{LR}_{I'_jA}$, if $f^{LR}_{I_j} > f^{LR}_{I'_j}$, then $\Theta^{LR}_{I_jA}$ is more likely to be true than $\Theta^{LR}_{I'_jA}$. Therefore, we test H_0 against $\Theta^{LR}_{I^0_jA}$ if $f^{LR}_{I^0_j} > \max\{f^{LR}_{I_j} : \forall I_j \neq I^0_j\}$, or equivalently if $p_{I^0_j}(f^{LR}_{I^0_j}) < \min\{p_{I_j}(f^{LR}_{I_j}) : \forall I_j \neq I^0_j\}$, for some I^0_j.

To make a choice of Θ^{LR}_{jA} between different j, one has to deal with the p-values rather than the realizations because now different F-distributions are involved. Fortunately, these p-values only depend on F-distributions as shown in Lemma 4.1 and can be easily computed by most statistical software. Let

$$p_j(f^{LR}_j) \overset{\text{def}}{=} \min\{p_{I_j}(f^{LR}_{I_j}) : \forall I_j\}. \tag{4.8}$$

For two different j and j', Θ^{LR}_{jA} is more likely to be true than $\Theta^{LR}_{j'A}$ if $p_j(f^{LR}_j) < p_{j'}(f^{LR}_{j'})$.

Now we are ready to partition the sample space. Let

$$S^{LR}_* = \{\text{all } p_{I_j}(f^{LR}_{I_j})\text{'s are different}\} \tag{4.9}$$

be the reduced sample space (note $P_\theta(S^{LR}_*) = 1$). Let

$$S^{LR}_{I_j} = \{p_{I_j}(f^{LR}_{I_j}) = \min\{p_{I_{j'}}(f^{LR}_{I_{j'}}) : \forall I_{j'} \subset \{1, \dots, p\}, \forall j' \in [0, p-1]\}\}. \tag{4.10}$$

Then all such $S^{LR}_{I_j}$'s form a partition of S^{LR}_*. The hypotheses of interest in this section are

$$H_0 : \beta_1 = \cdots = \beta_p = 0 \ \text{ vs } \ H_A : \theta \in \Theta^{LR}_{I_jA}, \ \text{ if } X \in S^{LR}_{I_j}, \forall I_j, \ \ \forall j \in [0, p-1]. \tag{4.11}$$

Thirdly, to determine an acceptance region within each of $S^{LR}_{I_j}$, first define a test statistic

$$F^{LR}_{I_j, j+1} = \min\left\{ \frac{\hat\beta_{I_{j+1}} \Sigma^{-1}_{I_{j+1}} \hat\beta'_{I_{j+1}} - \hat\beta_{I_j} \Sigma^{-1}_{I_j} \hat\beta'_{I_j}}{(\hat\beta_{I_j} \Sigma^{-1}_{I_j} \hat\beta'_{I_j} + SSE)/(n-p-1+j)} : \forall\, I_{j+1} \supset I_j \right\} \tag{4.12}$$

and a region

$$R^{LR}_{I_j, r} = \{F^{LR}_{I_j, j+1} > r\} \tag{4.13}$$

for a positive constant r. The set $R^{LR}_{I_j,r}$ converges to the empty set or the sample space when r goes to infinity or zero. Now let

$$R^{LR}_{I_j}(r^{LR}_{I_j}) = R^{LR}_{I_j, r^{LR}_{I_j}} \cap S^{LR}_{I_j} \tag{4.14}$$

for some constant $r^{LR}_{I_j} > 0$ so that

$$P_{\theta_0}(R^{LR}_{I_j}(r^{LR}_{I_j}) | S^{LR}_{I_j}) = \alpha. \tag{4.15}$$

The set $R_{I_j}^{LR}(r_{I_j}^{LR})$ is the acceptance region of level α within $S_{I_j}^{LR}$. On this set, $\{\beta_i : i \in I_j\}$ are likely to be zero due to $S_{I_j}^{LR}$, and $\{\beta_i : i \notin I_j\}$ are likely to be nonzero due to $R_{I_j, r_{I_j}^{LR}}^{LR}$. The construction of the acceptance regions is complete. Following Theorem 2.1, we have

Theorem 4.1. *The rejection region*

$$R^{LR}\left(= \bigcup_{j=0}^{p-1} \bigcup_{all\ \ I_j} R_{I_j}^{LR}(r_{I_j}^{LR})\right) \tag{4.16}$$

defines a level-α test for (4.11)*, as well as* (4.2).

Remark 4. There exists flexibility for the construction of $S_{I_j}^{LR}$'s. Suppose the researcher only wants to use a certain number of predictors in the model, say the number of predictors belongs to a predetermined set $G \subset \{1, \ldots, p\}$, typically a range. One can make this happen by changing (simplifying) the partition (4.10), correspondingly. For example, if a simple linear regression model is not an appropriate choice, then $G = \{2, \ldots, p\}$ and the partition (4.10) is replaced by

$$S_{I_j}^{LR} = \{p_{I_j}(f_{I_j}^{LR}) = \min\{p_{I_j}(f_{I_{j'}}^{LR}) : \forall I_{j'}, j' \in [0, p-2]\}\}, \tag{4.17}$$

and then any of the $S_{I_{p-1}}^{LR}$'s is an empty set. This becomes very important when the number of possible predictors, p, is large, because the number of subsets in the original partition is equal to $2^p - 1$, and comparing $2^p - 1$ of p-values is computationally extensive. However, the researcher needs a better understanding of the situation to choose G.

Example 2 (continued)**.** Assume model (4.1) applies to the data, and the goal is to identify all nonzero β_i's under this model with $p = 4$, $n = 13$ and a controlled type I error of level 0.05. For application purpose, we provide details step-by-step.

a) First identify the appropriate alternative $\Theta_{I_j A}^{LR}$. From the data set, compute $p_{I_j}(f_{I_j}^{LR})$, as shown in Table 1. We find that $p_{\{3,4\}}(f_{\{3,4\}}^{LR}) = 4.407 * 10^{-9}$ is the smallest p-value. Therefore, the data fall in

Table 1 The observed $f_{I_j}^{LR}$ and the p-values $p_{I_j}(f_{I_j}^{LR})$

I_3	{1,2,3}	{1,2,4}	{1,3,4}	{2,3,4}		
$f_{I_3}^{LR}$	22.799	4.403	21.961	12.603		
$p_{I_3}(f_{I_3}^{LR})$	0.000576	0.0598	0.000665	0.00455		
I_2	{1,2}	{1,3}	{1,4}	{2,3}	{2,4}	**{3,4}**
$f_{I_2}^{LR}$	72.267	10.628	27.685	176.628	6.066	**229.504**
$p_{I_2}(f_{I_2}^{LR})$	$1.134 * 10^{-6}$	0.00335	$8.377 * 10^{-5}$	$1.581 * 10^{-8}$	0.0188	$\mathbf{4.407 * 10^{-9}}$
I_1	{1}	{2}	{3}	{4}		
$f_{I_1}^{LR}$	107.375	157.266	166.832	166.345		
$p_{I_1}(f_{I_1}^{LR})$	$2.302 * 10^{-7}$	$4.312 * 10^{-8}$	$3.323 * 10^{-8}$	$3.366 * 10^{-8}$		
I_0	$\emptyset$					
$f_{I_0}^{LR}$	111.480					
$p_{I_0}(f_{I_0}^{LR})$	$4.756 * 10^{-7}$					

$S_{\{3,4\}}^{LR}$, and we test hypotheses:

$$H_0 : \beta_i = 0,\ \ \forall\, i = 1, \ldots, 4,\ \ \text{vs}\ \ H_A :\ \ \beta_1 \neq 0,\ \ \beta_2 \neq 0,\ \ \beta_3 = \beta_4 = 0,$$

within $S_{\{3,4\}}^{LR}$.

b) Compute the cutoff point $r_{\{3,4\}}^{LR}$. Although the cutoff point is uniquely defined under θ_0, its exact value is very difficult to compute analytically. So we calculate this by 5,000,000 simulations coded in Gauss 7.0 and obtain $r_{\{3,4\}}^{LR} = 8.035$ following two steps below:

- b1. In each simulation, we generate $\hat{\beta}_I$ for $I = \{1, 2, 3, 4\}$ following $N_4(0, \Sigma)$ and a SSE following χ^2 distribution with 8 degrees of freedom, and calculate $\{p_{I_j}(f_{I_j}^{LR})\}_{\forall I_j, j \in [0,3]}$(15 p-values). If the minimum p-value is achieved at $I_j = \{3, 4\}$, then the simulated sample is considered as a success, and the realization of $F_{\{3,4\},3}^{LR}$ in (4.12) is computed and recorded; otherwise the simulated sample is discarded.

- b2. Sort the set of $F^{LR}_{\{3,4\},3}$-values from all successful simulated samples. The upper 95th percentile of this sorted set is the simulated value of $r^{LR}_{\{3,4\}}$, which equals 8.035.

c) Compute the test statistic, $F^{LR}_{\{3,4\},3}$, given in (4.12), based on the observed data. Note that there are two sets of three elements $\{1,3,4\}$ and $\{2,3,4\}$ containing $\{3,4\}$. Thus

$$F^{LR}_{\{3,4\},3} = \min\{146.523, 208.583\} = 146.523,$$

much larger than $r^{LR}_{\{3,4\}}$. So we conclude that only X_1 and X_2 are in the model at level 0.05.

Remark 5. In order to obtain a conclusion, we only need to find which $S^{LR}_{I_j}$ contains the data set and the corresponding $r^{LR}_{I_j}$. So only a single number $r^{LR}_{I_j}$ for $I_j = \{3,4\}$, not the whole set of $r^{LR}_{I_j}$, is needed to draw the conclusion.

Compared with the traditional model selection methods (see details [13, p. 316]), the predictors included in the final model are shown below:

The proposed: X_1, X_2; forward selection: X_1, X_2, X_4; backward elimination: X_1, X_2; stepwise regression: X_1, X_2; all subsets (with two predictors): X_1, X_2.

The proposed test yields a similar result. However, the proposed test controls both the overall and conditional type I errors at 0.05 level, while the others do not. Also, the proposed test does not involve subjective choices on when to add or delete the predictors from the model.

5 Applications in multi-factor analysis

We apply the proposed testing procedure to two-way and three-way analysis of variance. For simplicity, balanced designs are assumed. The partition on the sample space is obtained using Method II) in Section 2.

5.1 Two-way anova

Assume a model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad 1 \leqslant i \leqslant a, \quad 1 \leqslant j \leqslant b, \quad 1 \leqslant k \leqslant n, \tag{5.1}$$

where μ is the fixed overall mean, α_i is either a fixed or random effect for factor A, the same assumption holds for factor B, $(\alpha\beta)_{ij}$ is either a fixed or random interaction between the two factors, and the ε_{ijk} are independent normal errors with a common unknown variance σ^2. The goal here is to detect whether there exist interactions between the two factors, A main effects and B main effects. Many authors, see [10, p. 848 and p. 1013] for example, suggest to test the existence of interactions first; and if no interactions, then test each factor separately for main effects. Thus up to three tests are needed. Here we can finish the analysis in a single step. The basic hypotheses are

$$H_0 : Y_{ijk} \sim \text{ i.i.d. } N(\mu, \sigma^2) \text{ (no any effect) vs } H_A : \text{ not so.} \tag{5.2}$$

First partition the above alternative as a union of the following four disjoint sets

$$\Theta_{1A}: \text{ there is interaction between two factors;} \tag{5.3}$$

$$\Theta_{2A}: \text{ no interaction, no } B \text{ main effect and there is } A \text{ main effect;} \tag{5.4}$$

$$\Theta_{3A}: \text{ no interaction, no } A \text{ main effect and there is } B \text{ main effect;} \tag{5.5}$$

$$\Theta_{4A}: \text{ no interaction, there are both } A \text{ and } B \text{ main effects.} \tag{5.6}$$

Next, to partition the sample space, let SSA, SSB, $SSAB$ and SSE be the sum of squares for A, B, AB and error, respectively. Let

$$F_1^{T2} = \frac{SSAB/((a-1)(b-1))}{SSE/(ab(n-1))}, \tag{5.7}$$

$$F_2^{T2} = \frac{SSA/(a-1)}{(SSE + SSAB + SSB)/(a(bn-1))}, \tag{5.8}$$

$$F_3^{T2} = \frac{SSB/(b-1)}{(SSE + SSAB + SSA)/(b(an-1)}, \tag{5.9}$$

$$F_4^{T2} = \frac{(SSA + SSB)/(a+b-2)}{(SSE + SSAB)/(abn-a-b+1)}, \tag{5.10}$$

which correspond to (5.3), (5.4), (5.5) and (5.6), respectively. Here the superscript $T2$ stands for two-way anova. Note that each F_i^{T2} follows an F-distribution under H_0 with the degrees of freedom given in the denominator in each numerator and denominator in each of the above ratios. Let f_i^{T2} be a realization of F_i^{T2}. For an observed f_i^{T2}, define a p-value for H_0 vs $H_{iA} : \Theta_{iA}$ as

$$p_i(f_i^{T2}) = P_{H_0}(F_i^{T2} \geqslant f_i^{T2}), \tag{5.11}$$

which is obtained by an F-distribution. Let

$$S_i^{T2} = \{p_i(f_i^{T2}) < \min\{p_j(f_j^{T2}) : j \neq i\}\}, \ \ \forall\, i = 1, 2, 3, 4. \tag{5.12}$$

The four sets S_i^{T2} form an almost surely partition of the sample space. Therefore, the hypotheses of interest in this subsection are

$$H_0 : Y_{ijk} \sim \ \text{i.i.d. } N(\mu, \sigma^2) \ \text{ vs } \ H_A : \ \Theta_{iA} \ \text{ is true if } \ X \in S_i^{T2}, \forall\, i = 1, \ldots, 4. \tag{5.13}$$

Finally, for $i = 1, 2, 3$, define an acceptance region in S_i^{T2} of the form

$$R_i^{T2} = \{F_i^{T2} > r_i^{T2}\} \cap S_i^{T2}, \tag{5.14}$$

while for $i = 4$, let

$$R_4^{T2} = \left\{ \frac{\min\{SSA/(a-1), SSB/(b-1)\}}{(SSAB + SSE)/(abn-a-b+1)} > r_4^{T2} \right\} \cap S_4^{T2}, \tag{5.15}$$

for some positive constant r_i^{T2} so that

$$P_{H_0}(R_i^{T2} | S_i^{T2}) = \alpha, \quad \forall\, i = 1, 2, 3, 4. \tag{5.16}$$

Following Theorem 2.1, we have

Theorem 5.1. *The rejection region*

$$R^{T2} \left(= \bigcup_{i=1}^{4} R_i^{T2} \right) \tag{5.17}$$

defines a level-α test for (5.13)*, as well as* (5.2)*.*

Remark 6. The traditional F-tests for factor main effects use either MSE or MSAB as the denominator depending on whether model (5.1) has fixed, mixed or random effects. However, the proposed test (5.17) is the same for all three cases.

Remark 7. A test of (5.17) can also be derived for an unbalanced two-way design by applying the generalized linear test approach to construct a partition for the sample space and construct a test. See a similar result in Section 4.

5.2 Three-way anova

It is well known that there may not exist an exact F-test for certain factor effects when the number of factors is at least three. For example, in a three-way anova with three random factors, A, B, C, no exact F-test is available to test each of the factor main effects. One solution is to use Satterthwaite's method

Table 2 The partitioned alternatives of H_A in (5.19)

$\Theta_{1A} = (ABC)^1$
$\Theta_{2A} = (ABC)^0 \cap (AB)^1 \cap (AC)^1 \cap (BC)^1$
$\Theta_{3A} = (ABC)^0 \cap (AB)^1 \cap (AC)^1 \cap (BC)^0$
$\Theta_{4A} = (ABC)^0 \cap (AB)^1 \cap (AC)^0 \cap (BC)^1$
$\Theta_{5A} = (ABC)^0 \cap (AB)^0 \cap (AC)^1 \cap (BC)^1$
$\Theta_{6A} = (ABC)^0 \cap (AB)^1 \cap (AC)^0 \cap (BC)^0 \cap C^1$
$\Theta_{7A} = (ABC)^0 \cap (AB)^1 \cap (AC)^0 \cap (BC)^0 \cap C^0$
$\Theta_{8A} = (ABC)^0 \cap (AB)^0 \cap (AC)^1 \cap (BC)^0 \cap B^1$
$\Theta_{9A} = (ABC)^0 \cap (AB)^0 \cap (AC)^1 \cap (BC)^0 \cap B^0$
$\Theta_{10A} = (ABC)^0 \cap (AB)^0 \cap (AC)^0 \cap (BC)^1 \cap A^1$
$\Theta_{11A} = (ABC)^0 \cap (AB)^0 \cap (AC)^0 \cap (BC)^1 \cap A^0$
$\Theta_{12A} = (ABC)^0 \cap (AB)^0 \cap (AC)^0 \cap (BC)^0 \cap A^1 \cap B^0 \cap C^0$
$\Theta_{13A} = (ABC)^0 \cap (AB)^0 \cap (AC)^0 \cap (BC)^0 \cap A^0 \cap B^1 \cap C^0$
$\Theta_{14A} = (ABC)^0 \cap (AB)^0 \cap (AC)^0 \cap (BC)^0 \cap A^0 \cap B^0 \cap C^1$
$\Theta_{15A} = (ABC)^0 \cap (AB)^0 \cap (AC)^0 \cap (BC)^0 \cap A^1 \cap B^1 \cap C^0$
$\Theta_{16A} = (ABC)^0 \cap (AB)^0 \cap (AC)^0 \cap (BC)^0 \cap A^1 \cap B^0 \cap C^1$
$\Theta_{17A} = (ABC)^0 \cap (AB)^0 \cap (AC)^0 \cap (BC)^0 \cap A^0 \cap B^1 \cap C^1$
$\Theta_{18A} = (ABC)^0 \cap (AB)^0 \cap (AC)^0 \cap (BC)^0 \cap A^1 \cap B^1 \cap C^1$

to derive an approximate F-test. Here we show in an example that approximation is not needed if the proposed method is applied. Assume a three-way anova model with three random factors A, B and C.

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}, \tag{5.18}$$

where all components, except for a constant μ, are independent normal random variables with mean 0 and variances: σ_α^2, σ_β^2, σ_γ^2, $\sigma_{\alpha\beta}^2$, $\sigma_{\alpha\gamma}^2$, $\sigma_{\beta\gamma}^2$, $\sigma_{\alpha\beta\gamma}^2$ and σ^2, respectively, and $1 \leqslant i \leqslant a, 1 \leqslant j \leqslant b, 1 \leqslant k \leqslant c, 1 \leqslant l \leqslant n$. No exact F-test exists for testing whether $\sigma_\alpha^2 = 0$ (i.e., no A main effect). Now the basic hypotheses are

$$H_0 : Y_{ijkl} \sim \text{ i.i.d. } N(\mu, \sigma^2) \text{ (no any effect) vs } H_A : \text{ not so.} \tag{5.19}$$

To partition the alternative, let J denote an effect and let J^1 denote that effect J exists and J^0 denote no such effect. Then H_A is the union of the disjoint sets Θ_{iA} given in Table 2. For example, Θ_{15A} says no any effects except for the A and B main effects. Similar to Section 5.1, we introduce an F-ratio for each Θ_{iA} and its corresponding p-value, construct an almost surely partition, $\{S_i^{T3}\}_{i=1}^{18}$, on the sample space, and construct an acceptance region R_i^{T3} of level α within each S_i^{T3}. We, for example, illustrate the process on Θ_{15A}. Let

$$F_{15}^{T3} = \frac{(SSA + SSB)/(a+b-2)}{(SSE + SSABC + SSAB + SSAC + SSBC + SSC)/(abcn - a - b + 1)}$$

and f_{15}^{T3} be its realization. Then the p-value is given by

$$p_{15}(f_{15}^{T3}) = P_{H_0}(F_{15}^{T3} \geqslant f_{15}^{T3}) = P(F_{a+b-2, abcn-a-b+1} \geqslant f_{15}^{T3}),$$

and

$$S_{15}^{T3} = \{p_{15}(f_{15}^{T3}) < \min\{p_i(f_i^{T3}) : 1 \leqslant i \leqslant 18, i \neq 15\}\},$$

provided that the $p_i(f_i^{T3})$'s are all computed. If the data fall in S_{15}^{T3}, then we test H_0 against $H_A : \Theta_{15A}$ within S_{15}^{T3}. Define

$$R_{15}^{T3} = \left\{ \frac{(abcn - a - b + 1)\min\{SSA/(a-1), SSB/(b-1)\}}{(SSE + SSABC + SSAB + SSAC + SSBC + SSC)} > r_{15}^{T3} \right\} \cap S_{15}^{T3},$$

where constant r_{15}^{T3} is determined by $P_{H_0}(R_{15}^{T3} | S_{15}^{T3}) = \alpha$. Then R_{15}^{T3} is an (exact) acceptance region of level α within S_{15}^{T3}. Following Theorem 2.1, we obtain a level α test with a rejection region $R^{T3} (= \bigcup_{i=1}^{18} R_i^{T3})$ for

$$H_0 : Y_{ijkl} \sim N(\mu, \sigma^2) \text{ vs } H_A : \ \Theta_{iA} \text{ is true if } X \in S_i^{T3}, \quad \forall\, i = 1, \ldots, 18. \tag{5.20}$$

6 Discussions

We develop a method that turns a multiple test problem into a single test problem by introducing a random alternative. This method generalizes the traditional testing procedure so that one can choose a claim out of more than two mutually exclusive claims, which typically provides a much more precise statement regarding the parameter. In fact, all testing problems with a large alternative hypothesis space may be reconsidered under the new setting. Some selected applications of the proposed are discussed in Sections 3, 4 and 5. Even for these classical problems, we still obtain more effective tests. One immediate implication is that the two-sided t-test, which is widely used in daily statistical practice, should be replaced by the proposed. Model selection for regression is also a very important problem in practice, and now one can make an objective selection with the type I error controlled at level α. Other applications can be found in one-way anova, orthogonal and nonorthogonal saturated designs, outliers detection, comparing several treatments with a control, change points and statistical quality control, etc.

The structure of the proposed test, listed in the end of Section 2, is simple and clear. One needs to partition the sample space based on the partitioned alternatives. We recommend Method II) in Section 2 to obtain a partition on the sample space since it is objective. However, one needs to make an effort to simplify the computation of p-values when a fine partition is involved. Within each set of the partition, the proposed test reduces to the traditional test. Thus all the techniques developed there to derive a test can be applied here.

When two level α tests are derived for the same set of H_0 and H_A, it is natural to compare their powers. Let R^T denote the rejection region of a traditional level α test for (1.5). It would be of interest to compare $P_\theta(R^T)$ with $P_\theta(R)$ for $\theta \in H_A$, where R is given in (2.3), even though R^T and R do not have exactly the same H_A(note that R deals with a random alternative). For some simple case, such as the t-test discussed in Section 3, the two powers are identical simply following Theorem 3.1. A general conclusion on the power comparison, however, is hard to obtain because the proposed test deals with a large class of hypotheses, and this deserves further study.

Acknowledgements This research was supported in part by US National Science Foundation (Grant No. DMS-0906858). The author wishes to thank two anonymous referees for their valuable comments which lead to the improvement of this paper, and Professor Dan Voss for proofreading the manuscript.

References

1 Anderson T W. An Introduction to Multivariate Statistical Analysis, 2nd ed. New York: Wiley, 1984
2 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B, 1995, 57: 289–300
3 Dunnett C W, Tamhane A C. A step-up multiple test procedure. J Amer Statist Assoc, 1992, 87: 162–170
4 Finner H, Strassburger K. The partitioning principle: a powerful tool in multiple decision theory. Ann Statist, 2002, 30: 1194–1213
5 Hald A. Statistical Theory with Engineering Applications. New York: Wiley, 1952
6 Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 1988, 75: 800–803
7 Hogg R V, Fisher D M, Randles R H. A two-sample adaptive distribution-free test. J Amer Statist Assoc, 1975, 70: 656–661
8 Hogg R V, Lenth R V. A review of some adaptive statistical techniques. Comm Statist Theory Methods, 1984, 13: 1551–1579
9 Holm S. A simple sequentially rejective multiple test procedure. Scand J Statist, 1979, 6: 65–70
10 Kunter M H, Nachtsheim C J, Neter J, et al. Applied Linear Statistical Models, 5th ed. Irwin: McGraw-Hill, 2005
11 Lehmann E L, Romano J P. Generalizations of the familywise error rate. Ann Statist, 2005, 33: 1138–1154
12 Marcus R, Peritz E, Gabriel K R. On closed testing procedures with special reference to ordered analysis of variance. Biometrika, 1976, 63: 655–660
13 Montgomery D C, Peck E A, Vining G G. Introduction to Linear Regression Analysis, 3rd ed. New York: Wiley, 2001
14 O'Gorman T W. Applied Adaptive Statistical Methods. Tests of Significance and Confidence Intervals. ASA-SIAM Series on Statistics and Applied Probability, 12. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), Alexandria, VA: American Statistical Association, 2004